M. Pittman 4/25/51

MEDICAL BASIC SCIENCES COURSE NO. 8-0-26
1950

- - - - - - - - - - -

MEDICAL STATISTICS

Herbert C. Batson, PhD.
Chief, Department of Biologic Products
Army Medical Department Research and Graduate School
Washington, D.C.

Captain Robert C. MacDuffee, M.C.
Captain Leon Eisenberg, M.C.

Experimentation

The nature of experimentation.—For the purpose of identifying the subject of this discussion, experimentation will be considered as that type of planned human endeavor by means of which new information is gained concerning the relationship of effect to cause. It is considered that, by definition, experimentation involves the attempt to control causes and the deliberate observation or measurement of effects as structurally related to those causes. The objective of such deliberate action is to obtain data or information from which, following evaluation and interpretation, knowledge may be gained. However, it is not meant to imply that in all experimentation causes actually are controlled successfully, observations or measurements are performed accurately, or even that the information gained is valid. Knowledge gained by experimentation, therefore, is not necessarily synonymous with truth.

The foregoing definition essentially excludes the natural philosophers from the category of scientific experimenters since they acquired information primarily by careful observation of naturally occurring phenomena over which they exercised no control. From the consideration of their compiled data, they attempted to deduce the underlying laws determining or governing the character or behavior of the subject of interest. The natural philosophers were the prototypes of the modern experimenters and it would be wasteful and presumptive to deny the value of their observations and deductions. Yet, it appears evident that if scientific knowledge had been acquired only by the means at their command, our present information would be severely limited in quantity, since information would be gained only by the concomitant condition of natural occurrence and competent observation, and limited in value or validity since correctness of interpretation would be dependent solely upon the acumen and deductive reasoning capability of the observer.

The limitation of quantity.—The scientifically curious early recognized, at least in part, the aforementioned limitations of the methods of the natural philosophers and it was inevitable that the impatient initiated change. The birth of deliberate experimentation involved only an extension of the existing methods,—the introduction of the technique of deliberate control of conditions or environmental factors. This permitted observation of the cause-effect relationship under combinations of conditions virtually unlimited in variety. The result of this change in methodology is known by all students of science. The accumulation of scientific information (at least of data) has reached such proportions that the magnitude of the body of organized knowledge is beyond the comprehension of most, and even the astute no longer can hope to achieve mastery and understanding of more than an extremely narrow segment of the whole.

The acquisition of scientific information no longer is a deductive adventure, the exclusive prerogative of the philosophers. It has become a profession,-highly organized, extremely competitive and, unfortunately, in too many instances, politically and economically ruthless. It is perhaps an inevitable, but none the less unfortunate, sociological anomaly that in science professional recognition and advancement frequently are meted out in general proportion to the volume of experimentation conducted and published by the researcher. This, in turn, has stimulated experimenters to produce greater and greater volumes of published research, often at the expense of quality and completeness and even, at times, with little regard to the real need for the information.

Tools of the professional researcher include, among others; apparati from simple tools to precision instruments of almost unbelievable complexity and sensitivity; common and rare chemicals (stable or unstable) of almost any desired purity; experimental animals, plants and microorganisms of a number and variety restricted only by the limitations of natural evolution and the ingenuity of geneticists; and the body of information, concepts and techniques provided by his colleagues. Thus, the first limitation to the acquisition of scientific knowledge by the methods of the natural philosophers, the limitation of quantity, to a large measure has been overcome by the introduction of the method of deliberate experimentation, sociological change, professional ambition and technological advance.

The limitation of value.--The second limitation of the methods of the natural philosophers, the limitation in value or quality, has not fared as well. It is not meant to imply that progress has not been made; such would be ridiculous. Yet the quality of information gained by scientific researchers, to a considerable extent, is dependent upon different factors than is quantity. The quality of the information is not dependent alone upon the precision or exactitude with which a cause-effect relationship is observed or measured. The attributes or effects which the experimenter observes or measures may be partly or wholly unrelated to the assumed causes and the experiment may have been so designed and performed that it is impossible to measure or estimate the true effects of causes or conditions in which the experimenter is interested. The supposed facts and assumptions accepted by the researcher in designing and interpreting his experiments actually may be erroneous. The restrictions inadvertently placed on the experimental complex by employing the ultimate in technical equipment may result in observations being made under such limited conditions that the observed effects would have no counterpart under conditions of natural occurrence and, yet, such results often are assumed to have broad or even universal application.

Furthermore, and probably most important of all, regardless of the precision of individual observations or measurements, the collected data must be interpreted and the scientific importance or significance of the findings must be assessed by the investigator. The final conclusions drawn by the researcher thus are dependent directly upon his judgment and experience and, perhaps, even his intuition. As a consequence, the quality of the information obtained by the investigator is limited to a degree proportional to his capability to interpret correctly and to assess the results of his own experiments and those of his professional colleagues whenever he wishes to incorporate their findings or conclusions as assumptions or supposed facts in the performance or evaluation of his own work. The latter requirement, evaluation and interpretation of the work of others, constitutes one of the major problems confronting the modern experimenter.

Scientific specialization and the quality of information.—Scientific research has become so complex and so demanding of special technical knowledge and skills, that the individual worker has been forced to limit his efforts to an ever narrowing segment of inquiry. Through this extreme specialization, he has been denied the benefits of study and personal experience in other fields and thus has become more and more dependent upon his colleagues. The problem of ever increasing volume and complexity of technical information and ever narrowing scope of individual competence has been met, in part, by the establishment of investigative groups or teams made up of individuals or subgroups expert in the fields of specialization essential to the total effort. This is a logical and inevitable sequence in the evolution of a complex society. The accomplishments of such investigative groups have been too dramatic to permit denial of the merits of such an approach. Yet to consider that the basic problem, the limitation of the quality of information, has been solved thereby, is to ignore the facts and to misinterpret the nature of the limitation. Actually, the result of joint effort in scientific research has been mainly a further increase in the volume of information. The limitation of quality has remained; in fact it has been enhanced. The incapability of the individual to evaluate or to interpret segments of the information still exists. In addition, evaluation and interpretation of the whole of the information poses a task of even greater magnitude; a task often transcending the capability of any but the most highly proficient,—the true scientific generalists.

The need for assistance in the evaluation of data.—The foregoing implies that man, with the aid of modern technology, now is or soon will be able to collect scientific information more rapidly than it can be evaluated or interpreted, or more complex than he can comprehend. Blind acceptance of such an attitude would be defeatist and inimicable to mans' best interests. It is better to consider that full comprehension is attainable and that what is needed is the devel-

opment of techniques in methodology, and discipline in mental processes, which will enhance the efficiency and capacity of understanding. If man, because of his innate mental limitations, is unable successfully to interpret and comprehend scientific information because of its volume and complexity, he needs assistance. Science and technology have proved capable of providing highly efficient aids for the collection of information; likewise they should be capable of providing aids for the evaluation and interpretation of information. In other words, science and technology should be capable of providing efficient methods for evaluating data in an impersonal manner, for extracting conclusions from compiled data, and for assessing the validity of such conclusions—the validity being uniquely determined by the data and free from the personal bias of the investigator.

Methods such as the foregoing should soon greatly enrich contemporary scientific knowledge since they should permit extraction of a maximum of valid information from experimental data with rejection of the nonvalid, and should even point the way to the collection of further data with a minimum of waste, i.e., with a maximum of valid information per unit of effort. Such techniques should serve another purpose as well; they should permit, at least within limits, the reinspection of previously accepted information. This, no doubt, would result in the rejection of many erroneous scientific hypotheses and assumptions which have been accepted as facts and which, by their very formulation and continued acceptance, perpetuate the accumulation and acceptance of additional nonvalid concepts. The time is not yet here, and may never come, when evaluation, interpretation, and extraction of valid conclusions from data can be accomplished as a purely mechanical process. Nevertheless, great progress has been made in this regard; more progress than is realized by most workers in certain fields of scientific endeavor. These aids to the collection, evaluation and interpretation of data consist of modern concepts and techniques in statistics and biometrics[2] as employed in the design and analysis of experiments.

Statistics and Experimentation

Statistics as an aid to research.—Statistical methods, as an aid to experimentation, have been developed and employed extensively in agriculture and in certain other fields of scientific research, but only to a very limited extent in the basic medical sciences. Development of efficient statistical methods applicable to the latter sciences is the responsibility of the professional statisticians

---

[2] The science of statistics as applied to biological observations properly is designated as biometrics. However, for the sake of simplicity, the term statistics will be used throughout this Introduction.

and no doubt eventually will be forthcoming. In the interim, the worker in experimental medicine can benefit greatly by employing in his research, statistical methods developed primarily for application in a different field of inquiry. Such methods may be lacking in efficiency or, in specific instances, may even be inappropriate for application to medical data. These limitations need not occasion undue concern as they are greatly outweighed by the benefits to be derived.

It is unfortunate that in the course of formal training most medical researchers receive no more than the most superficial, cursory introduction to statistical methods and to their most fruitful field of application, the design of experiments. To most, statistics means only "tests of significance,"—a prosaic, essentially mechanical procedure whereby, with some mysterious but rather uninteresting mathematical manipulation, conclusions drawn from data can be labeled as "significant" or "not significant" with the exact connotation of either not quite understood by the researcher and the choice of labels, he frequently suspects, often dependent upon the malicious caprice of the statistician. Nothing could be farther from the truth than the foregoing. Statistics is essentially a way of thinking and, as such, constitutes the life blood of experimentation. Experimentation, to the true researcher, is an adventure—an excursion into the unknown; the rewards are the thrill of discovery—elucidation of the truth. The design of an experiment is the formal plan of attack. With the aid of modern statistics and foregained preliminary information, experiments often can be so designed that the chances of obtaining definitive results can be predetermined, the reliability of the results obtained can be assessed, the effects of multiple factors can be determined individually or in combinations with an actual economy in total effort, and the results of further similar experiments can be predicted with a minimum of uncertainty.

It must be emphasized that such benefits accrue primarily when statistical concepts and procedures are employed while planning and conducting experiments and that they should be utilized throughout an investigation. Unless the application of statistical procedures to the final data is taken into consideration when a problem is first planned, valid analysis frequently is impossible.

Use of statistics by the nonmathematician.—Statistics and formal experimental designs are mathematical in origin. Accordingly, mastery of mathematics by those who would use these techniques would be advantageous. Yet, to deny use of these methods to those who are incapable of full understanding of the mathematics involved would be no more justifiable than it would be to deny the use of a photoelectric colorimeter to the chemist incompetent in optics and electronics.

What characteristics and capabilities are required of the researcher who wishes to employ statistics as an integral part of his

experimentation?--A reasonable degree of proficiency in arithmetic and simple algebra, training and experience adequate to insure competent scientific judgment in his chosen field of research, the ability to think clearly, intuitive common sense sufficient to enlist professional aid if required, and the sustained honest desire to acquire understanding and proficiency in statistical methods. A sincere desire to learn statistics primarily is the result of personal conviction that the rewards of such an accomplishment are sufficiently attractive to justify the effort.

Why should the experimenter impose upon himself the added burden of acquiring understanding and proficiency in an additional field of science? Why should he not merely enlist the aid of the professional statistical consultant for analysis of his results?--Statistics is a powerful and useful tool but it should not be forgotten that the true information is intrinsic to the data and that the net worth of the individual observations is determined by the appropriateness of the conditions under which they are made. This is determined primarily by the scientific acumen and experience of the investigator. The statistician and the experimenter are mutually dependent throughout an investigation and research approaches its ultimate efficiency when they are one and the same individual.

Plan and Scope

It is planned in this Introduction to Statistical Methods to present certain basic statistical concepts and procedures and to show, by examples, how they may be employed in actual experimental investigations. It is beyond the scope of this Introduction to present examples of more than the most widely useful statistical procedures. Additional methods, suitable for use in many special situations, are available.

The particular statistical procedure applicable in any situation is dependent upon the type of the experiment conducted, the general nature of the data obtained, and the specific questions it is desired to answer. When the student becomes familiar with the different statistical techniques, he should experience little difficulty in selecting the appropriate procedure.

Classical vs. Modern Experimentation

The primary functions of statistics are to aid, first, in the extraction of information from data and, second, in the determination of the validity or reliability of such information. The total amount and true validity of information in any collection of observations are dependent upon these data and the manner in which they were obtained. This became increasingly apparent as statistical methods for analyzing data were developed in certain fields of research and, as a consequence, in those fields more and more attention has been directed towards improvement in the experimental procedure itself. The net result is the present existence of two general types of experimental procedures, unlike in methodology and based upon different concepts. For the purpose of further discussion (but without any necessary connotation of relative merit), those experimental procedures which have evolved independently of statistical considerations may be designated as classical, while those developed hand in hand with statistics may be designated as modern.

Classical experimentation.—The classical experimental approach is characterized by the attempt to determine the effect of a single condition or factor in such a manner that any observed effect on the experimental subject(s) may be ascribed to the single possible source. The procedure involves the variation of only one factor (such as the employment of different levels) in any single experiment, all other conditions or factors being held constant. Each factor of interest then is varied in turn while all remaining factors are held constant.

The classical experimental method is based (perhaps unwittingly) upon two concepts: first, that the effect of variation of a single factor is independent of concomitant variation in other factors; and, second, that the purpose of controlling experimental conditions (factors) is the minimization of error. Decision as to the dependability or trustworthiness of experimental information obtained by the classical procedure generally is based upon consideration of the size of the group of individuals studied, or the number of determinations made, under rigidly controlled conditions.

Modern experimentation.—Modern experimental procedures, as developed primarily in the science of agriculture under the leadership of R. A. Fisher, differ markedly from the preceding. First, as far as is practicable, those factors which are expected to affect the outcome are allowed to vary concomitantly within known limits within the individual experiment. This action is based upon the now recognized fact that the effect of variation in one factor frequently is dependent greatly upon concomitant variation in another factor or group of factors. This mutual dependence of the effect of factors commonly is referred to as interaction or discrepance and results from the failure of one

factor to exert a constant independent effect when associated with variation in another factor. Second, in the modern experimental procedure, more emphasis is placed upon accurate estimation of error than is placed upon its minimization. This constitutes a change in the concept of error from an association of the term with mistakes or inaccuracies in observations and measurements to a concept of normally existing variation in the experimental subjects or material independent of the known factors of experimental interest.

In modern experimentation, decision as to the dependability, trustworthiness or "significance" of an observed effect is not based upon the size of the experimental group employed so much as upon the frequency or repetition of observation of the effect. Thus, in modern experimentation, a given effect repeatedly obtained in 10 small experimental groups generally would be considered more meaningful than would a proportional or even greater effect obtained in a single group 10 times as large.

Statistical Terms and Concepts

The fundamental concepts and terminology (including symbols) of statistics are new to many engaged in research. Familiarity with the basic terminology, of course, is essential for employment of statistical computational procedures but the true value of such knowledge extends much farther. In statistics, as in most sciences, familiarity with the basic terminology, provided such familiarity constitutes appreciation and not just rote memorization, in itself assures a certain level of understanding of the subject and the limits of its application. Such understanding, even though it be more nearly akin to intuition than to mathematical comprehension, frequently is sufficient to permit competent utilization of statistical methods as an integral part of the experimental endeavor.

In accordance with the foregoing, the manner of presentation of the fundamental statistical terms and concepts, which follows, has been selected with a view to promoting understanding.

Population.—As used in statistics, the population is the total group of individuals of a like nature. At least, the members of the group are alike or similar in regard to certain distinguishing characteristics or attributes. The term is applied to groups of either animate or inanimate individuals, such as humans, animals, plants or articles; and even to properties, characteristics, observations, measurements, etc. Examples of populations dealt with in medical research include: heights of humans, gains in weight of animals on certain diets, assays of the vitamin content of foodstuffs, susceptibilities of groups of humans or animals to infection, etc. Statistical computational procedures are based on the assumption that the population of interest is infinite. In actual practice, these methods, which by derivation are strictly applicable only to studying and characterizing infinite populations, are applied to populations definitely limited in size. This is a source of error which, on occasion, may be serious.

Sample.—A sample is a finite group of individuals (entities or values) selected or drawn from a population and which supposedly is representative of the population. To insure that samples are truly representative of the population of interest (i.e., unbiased) they should be random samples. Random samples are those obtained in such a manner that each member of the population has equal opportunity of being included. In practice it frequently is essentially impossible to obtain true random samples and, as a consequence, experimental data commonly are biased to some degree.

Parameter.—A parameter is a true value of some characteristic of the population, such as the true average. Except in rare and unusual instances, parameters or true population values are unknown.

Statistic.—A statistic is an estimate, derived from a sample, of a true population value or parameter.

Statistical inference.---The estimation of the true characteristics (parameters) of a population from information derived from a sample (statistics) is known as statistical inference.

Probability.---Probability is an expression of the likelihood of occurrence. Probability can be expressed as a proportion of the whole (0 to 1), as a percentage, or as odds. Thus, if in a group of 100 individuals there were 5 with a certain attribute, the probability of selecting one of these by pure chance alone could be expressed as the proportion 0.05, as 5%, or as odds of 1 in 20.

Variables.---In statistical terminology, a variable is any quantity denoting condition, occurrence or effect which can take a succession of values. In experimental work we are concerned with two main types of variables which differ fundamentally:

> Independent variables: These are the known values or conditions, chosen or imposed by the investigator, under which experimental observations are to be made. Independent variables commonly are referred to as factors, levels, or "treatments." In computational procedures values of independent variables most frequently are denoted as values of X.
>
> Dependent variables: These are the actual experimentally determined values, obtained by the experimenter as measurements or enumerations, and frequently referred to as "individuals" or "observations." The absolute value of any individual (observation) supposedly is dependent, at least in part, upon the particular combination of independent variables (factors) employed. In computational procedures values of dependent variables most frequently are denoted as values of Y.

Measurement and characterization of population samples.--- While each observation in a series individually conveys a certain amount of information, data usually are most reliable when the series or group is considered as a whole. Then, the behaviour of groups subjected to different experimental conditions (factors or levels) can be compared to permit decision as to the effect of the factor of interest. Thus, it is essential that certain characteristic properties of series of values (samples) be determined in order that the direct comparison of groups can be accomplished. The following are useful characteristics of such series of values and are employed widely in statistical analyses:

> Individual or observation: A numerical expression or entry connoting a value either in enumeration or in measurement.

Array or sequence: A series of numbers (observations or individuals). The number of entries usually is designated by n.

Total or summation: The arithmetic sum. This quantity usually is designated as Σ although some modern writers prefer to use the symbol S. For example, ΣY symbolizes the total or sum of all values of Y.

Mean: The mean is the arithmetic average of a series of values and usually is designated by placing a bar over the symbol. For example,

$$\bar{Y} = \frac{\Sigma Y}{n}$$

Range: The arithmetic difference between the largest and smallest entries in the sequence. The range usually is designated by w. (This symbol frequently is confused as w also is used to denote the weight or relative importance of a value.)

Deviation from the mean: This is the arithmetic difference between an individual value and the mean of all values in the sequence. These deviate values often are denoted by lower case letters. Thus, $Y-\bar{Y}$ and y are synonymous. In any sequence the summation of the deviations of the individual values from the mean must equal 0.

Degrees of freedom: The term, degrees of freedom, denotes the number of individual values in a sequence or group which can be varied independently and concomitantly without necessitating any change in the prescribed value of any other particular member of the group. Stated in another way, the degrees of freedom associated with any series of values numerically are equal to the number of members in the series minus the number of restrictions placed upon the series. Thus, a sequence of n independent values has n degrees of freedom. If the values are summed, however, one restriction has been placed on the sequence and the degrees of freedom are reduced to n-1. The concept of degrees of freedom is particularly important as each degree of freedom represents an independent comparison that can be made from the observations. Degrees of freedom are represented by D.F.

Variance: The variance or mean square is a particularly useful statistic used to express the degree of variability or scatter of a sequence of values. By definition, variance is the sum of the squares of the deviations of the individual values from the mean, divided by the degrees of freedom. The variance of a population is represented symbolically as σ^2 while an estimate of variance obtained from a sample is represented as s^2.

Thus, variance is expressed symbolically as

$$s_Y^2 = \frac{\Sigma (Y-\bar{Y})^2}{n-1} \text{ ; or } = \frac{\Sigma y^2}{n-1}$$

Standard deviation: The standard deviation is the statistic most commonly employed to denote the degree of variability of a sequence of values. It is computed as the square root of variance and usually is represented as σ or s when referring to population and sample values, respectively.

Standard deviation of a mean: When samples of any size, say n, are drawn at random from a population, the mean values of the several samples are found to be much more closely grouped around the population mean than are the original individual values. The standard deviation of the mean of a sample is computed as

$$s_{\bar{Y}} = s/\sqrt{n}.$$

Calculation of variance and the standard deviation: The nature of the variance and the standard deviation are more readily apparent from consideration of the procedure used in computation of these statistics. The following example is presented both symbolically and with actual numerical values to serve as a guide for such computations.

	Values of Y	Deviations	Squared deviations
	$Y_1 = 2$	$Y_1-\bar{Y} = y_1 = 2-4 = -2$	$y_1^2 = (-2)^2 = 4$
	$Y_2 = 3$	$Y_2-\bar{Y} = y_2 = 3-4 = -1$	$y_2^2 = (-1)^2 = 1$
n = 4	$Y_3 = 5$	$Y_3-\bar{Y} = y_3 = 5-4 = +1$	$y_3^2 = (+1)^2 = 1$
	$Y_4 = 6$	$Y_4-\bar{Y} = y_4 = 6-4 = +2$	$y_4^2 = (+2)^2 = 4$
Total:	$\Sigma Y = 16$	$\Sigma(Y-\bar{Y}) = \Sigma y = 0$	$\Sigma y^2 = 10$

Mean: $\bar{Y} = \frac{\Sigma Y}{n} = \frac{16}{4} = 4$

Degrees of freedom: (D.F.) $= n-1 = 4-1 = 3$

Variance: $(s^2) = \frac{\Sigma y^2}{n-1} = \frac{10}{3} = 3.33$

Standard deviation (s) $= \sqrt{s^2} = \sqrt{3.33} = 1.82$

Standard deviation of the mean $(s_{\bar{Y}}) = s/\sqrt{n} = 1.82/2 = 0.91$

In actual practice, a short cut method for calculation of

variance is available as follows:

$$s^2 = \frac{\Sigma Y^2 - \frac{(\Sigma Y)^2}{n}}{n-1}$$

Substituting the appropriate values from the foregoing example,

$$s^2 = \frac{(2^2 + 3^2 + 5^2 + 6^2) - \frac{16^2}{4}}{3} = 3.33$$

and

$$s = \sqrt{3.33} = 1.82.$$

This method will be found very useful whenever dealing with long series of values or large numbers. The quantity $\frac{(\Sigma Y)^2}{n}$ in the above formula is known as the correction factor (usually represented as C.F.) and obviates the calculation of the deviation of each value from the mean. Similar correction factors are employed in many statistical computations.

Distributions.--Whenever a series of observations or measurements are obtained, even though such measurements are made on the same experimental subject or article, the successive values will not be identical. Such variation may result from innate differences in experimental subjects, known or unknown environmental and experimental factors, and actual errors in measurement. When large numbers of measurements are available and the frequency of occurrence of values are plotted against the degree of deviation from the mean value, it is found that the distribution or frequency curve takes a characteristic shape. Most experimental data follow the Gaussian or normal distribution quite closely. In some instances it is necessary to transform the numerical values to logarithms or some other function in order to approximate the normal distribution.

The normal distribution: The normal distribution is completely defined by two parameters, the mean and the standard deviation. Typically, the two parameters are independent, i.e., the dispersion or scatter of the observations (standard deviation or variance) is independent of the magnitude of the mean value of the attribute or characteristic of interest. The normal distribution can be illustrated as a symmetrical bell shaped curve as shown in Figure 1.

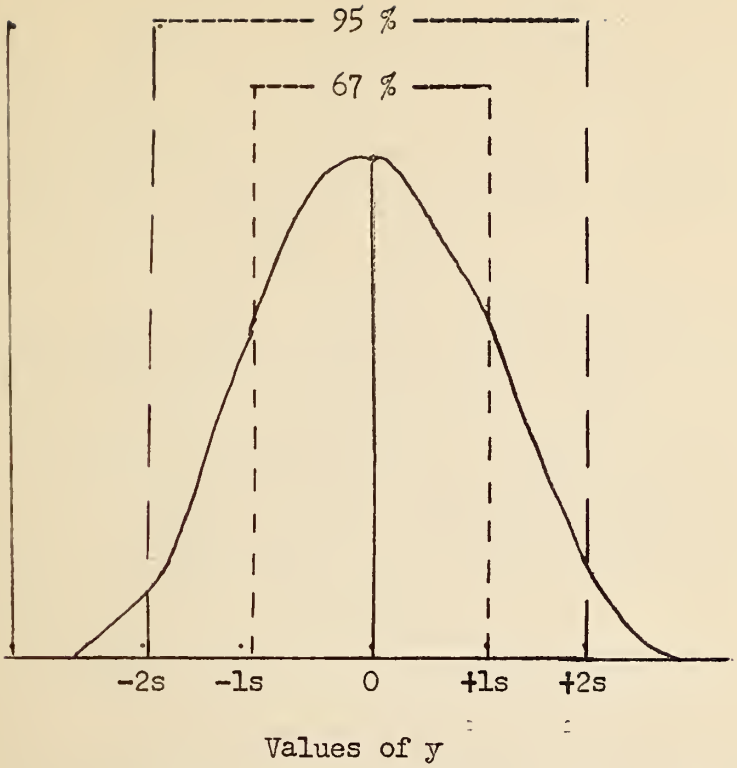


Figure 1. The normal distribution or normal curve of error.

It will be noted that there are many more individuals having values close to the mean value than there are those having values much greater or less than the mean. When the standard deviation of the individual values is calculated and unit standard deviations are marked off on the base line, the area of the curve between the ordinates erected at -1s and +1s constitutes approximately 67% of the total area. Likewise, the area between -2s and +2s constitutes approximately 95% of the total area. It follows from the foregoing, that the probability of selecting at random from a normal population an individual differing from the mean by a quantity d, becomes less as the ratio $\frac{d}{s}$ increases. This is the basis of the "t" test which will be described later. Thus, the probability of selecting at random an individual value (Y) differing from the mean (Y) by more than 2s is less than 0.05. In other words, we would expect to obtain such a value less than 5% of the time.

The binomial distribution: Certain types of attributes of particular interest to researchers in the biological sciences are not distributed normally. The occurrences of some event,

such as response or nonresponse, in successive samples of equal size, are not distributed normally but follow the binomial distribution. If there is a constant probability p of occurrence of the event, and a probability of nonoccurrence of q = 1-p, in samples of size k, the expected relative frequency of 0,1, 2,3, etc., occurrences in groups, is given by the successive terms of the binomial expansion of $(p + q)^k$. The variance of this distribution is kqp and thus is directly dependent upon the mean, being greatest when p = 0.5.

The Poisson distribution: The Poisson distribution is, in fact, the limiting form of the binomial in which p is very small. Thus, this distribution is of interest in relation to the distribution of events of rare occurrence. The Poisson distribution is characterized completely by a single parameter, the mean. The variance equals the mean. If the mean frequency of occurrence of an event is m, in a series of n samples, and the distribution truly is Poisson, the expected frequency of samples in which the event will occur 0, 1, 2, etc., times is given by the successive terms of the sequence—

$$n/e^m,\ mn/e^m,\ m^2n/2e^m,\ m^3n/(2)(3)e^m,\ m^4n/(2)(3)(4)e^m,\ \text{---}$$

The Poisson distribution is of special interest to bacteriologists since the distribution of organisms in dilute suspensions, such as are employed in preparing pour plates, is of this type.

Error.--All experimentally determined values are subject to error but, as used in statistics, the term error does not refer to specific mistakes as such. Rather, error refers to variations in individual values resulting from unknown causes. Error then connotes experimental uncertainty or discrepancy resulting from the vagaries of sampling and from failure to control, measure or recognize all causative factors. Such error is of a random nature. In larger and larger samples the mean value of random errors approaches 0 and their distribution is normal. In properly designed experiments, random errors are estimable directly even though their source is unknown.

Random error should not be confused with bias which is not estimable. Bias, when existent, is constant, i.e., its mean value is other than 0. Bias results from faulty assumptions, faulty experimental procedures, or faulty experimental design.

Significance.--As used in statistics, significance in reality is a statement of the probability that an observed effect or difference is a true effect or difference and not merely a chance occurrence. When, through performance of a formal test of significance, it is ascertained that it is improbable that an observed

effect or difference could have occurred by chance alone, the effect is designated as being statistically significant.

Designation of a given effect as being significant or non-significant depends, of course, on the level of probability of chance occurrence employed. Probability levels of chance occurrence of 0.05, 0.01 and 0.001 (corresponding to odds of 1 in 20, 1 in 100 and 1 in 1000) commonly are used in analysis and interpretation of experimental data. Observed effects whose estimated probability of chance occurrence correspond to (or are even less than) the foregoing are designated as being significant at the 5%, 1% and 0.1% levels, respectively.

The fact that an effect does not prove to be statistically significant is no proof that it does not exist. Lack of significance merely means that the effect could have resulted through chance alone.

Statistical significance should not be confused with experimental significance. An effect which statistically is significant is one which probably is not of chance occurrence; the effect, in itself, may not be of any real interest or importance.

Hypotheses (the null hypothesis).--Analysis and interpretation of experimental data always involves testing some (frequently more than one) hypothesis. Usually, in experimentation, it is desired to determine whether or not experimental subjects respond differently to various treatments or conditions, or if individuals from two or more groups differ in regard to some characteristic or attribute of interest to the investigator. Regardless of the nature of the experiment, when the findings are to be interpreted, various questions are posed, even though perhaps unconsciously.

The nature and identity of specific questions to be answered are determined uniquely by the design of the experiment and, in essence, answering these questions involves the employment of appropriate tests of significance. Yet the questions are not answered directly. Rather, hypotheses are set up,--hypotheses that the observed effects or differences are not true effects or differences, but were obtained merely by chance. These are null hypotheses. Formal tests of significance then are employed to determine the probability of obtaining results such as those observed if the null hypothesis indeed is correct. If the occurrence of the observed results by chance alone is found to be improbable, the null hypothesis is rejected and the alternate hypothesis, that a true effect or difference in fact exists, is accepted.

Computational Procedures and Tests of Significance

The Chi-square (X^2) Test

The Chi-square test, in general, is the statistical test appropriate for determining whether or not an observed frequency of occurrence of an event differs significantly from that expected in accordance with some hypothesis. Thus, this statistic is used primarily with data pertaining to frequency of occurrence rather than to degree; with enumeration rather than with measurement data or with discrete rather than with continuous variables. Chi-square is defined as follows:

$$\text{Chi-square} = \Sigma\left[\frac{(O-E)^2}{E}\right]$$

where Σ denotes the sum of all values; O, the observed frequency of occurrence; and, E, the expected frequency of occurrence in accordance with the hypothesis.

Chi-square is somewhat unreliable where applied to small numbers. In general, no class interval in which the expectancy of occurrence is less than 5 should be used alone but should be combined with an adjoining interval. Furthermore, Chi-square is a continuous variable but the data to which it is applied is discontinuous, i.e., they take successive values differing by whole units. Unless the numbers are large (100 or more) a more reliable estimate of Chi-square is obtained by making a correction for continuity by subtracting 0.5 from all values exceeding expectancy and adding 0.5 to all values less than expectancy.

The 1 x n table.—Chi-square is a useful statistic for testing an hypothesis of independence of occurrence of certain attributes in population samples. This application of the Chi-square test can be illustrated with data on the distribution of Rh blood types in a White population as reported by Wiener, Davidsohn and Potter (1). One thousand individuals were tested for the specific types Rh_0, Rh_1, Rh_2, and Rh_1Rh_2. The observed distribution of these types was reported as follows (Table 1):

Table 1. Distribution of Rh types in a White population. Data adapted from observed findings reported by Wiener, Davidsohn and Potter (1).

Rh type	Occurrence	
	Observed	Expected
Rh_0	26	215
Rh_1	541	215
Rh_2	128	215
Rh_1Rh_2	164	215
Total	859	860

In accordance with the null hypothesis, as previously discussed, it is assumed that there is no true difference in occurrence of the four Rh types, i.e., that the expected frequency of each is 215 (in round numbers; actual expectancy is 214 3/4). Chi-square then is computed as

$$\text{Chi-square} = \frac{(26-215)^2}{215} + \frac{(541-215)^2}{215} + \frac{(128-215)^2}{215} + \frac{(164-215)^2}{215}$$

$$= \frac{(-189)^2 + (326)^2 + (-87)^2 + (-51)^2}{215}$$

$$= 707.75$$

The value of Chi-square required for significance, i.e., to establish the improbability of chance occurrence of the observed results, is dependent upon the associated degrees of freedom. In this example Chi-square has three degrees of freedom. There are four values in the sequence but one restriction has been imposed, viz., the individual values were totalled to permit calculation of the expected values.

Referring to the table of Chi-square values (Table I, appendix A) it is found that, with three degrees of freedom, a value of Chi-square as large as 16.27 will arise by chance alone only once in 1000 times if there is no true difference in the frequency of occurrence of four attributes (in this case, Rh types). Clearly then, a value of Chi-square as large as 707.75 occurring as the result of chance alone is essentially inconceivable. Accordingly, the null hypothesis is rejected and it is concluded that the four Rh types do not occur with equal frequency in the White population. The unequal frequency of occurrence exceeds the 0.1% level of significance or, in statistical symbols.—

$$\text{Chi-square}_{[3]} = 707.75^{***}$$

where the number in brackets denotes the degrees of freedom, and the three asteriks[3] denote the level of significance equalled or exceeded.

The 2 x n table.—It frequently is desired to determine if the occurrence of certain characteristics or attributes is the same (i.e., not significantly different) in two population samples. An example of this type of problem also is afforded by the data in the report referred to in the preceding section (1). In that study the occurrence of the four Rh types was determined in a Negro as well as in a White population. The observed distribution of the types (underlined numbers) in the two populations is presented in Table 2.

---

[3] The 5% level of significance is denoted by *; the 1% level by **.

Table 2. Distribution of Rh types in the White and Negro Populations. Data adapted from observed findings reported by Wiener, Davidsohn and Potter (1).

Rh type	Occurrence				Totals
	Whites		Negroes		
Rh_0	a	26 (96)	b	92 (22)	118
Rh_1	c	541 (476)	d	45 (110)	586
Rh_2	e	128 (145)	f	50 (33)	178
Rh_1Rh_2	g	164 (143)	h	12 (33)	176
Totals		859 (860)		199 (198)	1058

Under the assumption (null hypothesis) that the distribution of the Rh types is the same in the two populations, the expected values (numbers in parentheses) are calculated (in round numbers) for each cell of the table as follows:

Cell	Expected value
a	$\frac{118 \times 859}{1058} = 96$
b	$118-96 = 22$
c	$\frac{586 \times 859}{1058} = 476$
d	$586-476 = 110$
e	$\frac{178 \times 859}{1058} = 145$
f	$178-145 = 33$
g	$\frac{176 \times 859}{1058} = 143$
h	$176-143 = 33$

The contribution to Chi-square of each cell in the table then is computed in accordance with the formula and the eight values are summed.

Cell	Contribution to Chi-square
a	$\frac{(26-96)^2}{96} = 51.04$
b	$\frac{(92-22)^2}{22} = 222.73$
c	$\frac{(541-476)^2}{476} = 8.88$
d	$\frac{(45-110)^2}{110} = 38.41$
e	$\frac{(128-145)^2}{145} = 1.99$
f	$\frac{(50-33)^2}{33} = 8.76$
g	$\frac{(164-143)^2}{143} = 3.08$
h	$\frac{(12-33)^2}{33} = 13.36$
Total Chi-square	$= 348.25$

It will be noted that the expected values for cells b, d and f were fixed by the computed values for cells a, c and e and the corresponding marginal totals for Rh types. Likewise, the values for cells g and h are fixed by the sums of values a, c and e, and b, d and f, respectively, and the marginal totals for the White and Negro populations. Accordingly, there are only three degrees of freedom associated with the total Chi-square and again a value of 16.27 would be required for significance at the 0.1% level. Therefore, the determined value of 348.25 again is incontestably significant.

The 2 x 2 or fourfold table.—In the analysis of experimental findings there are many occasions to apply a Chi-square test for significance to a 2 x 2 or contingency table which is, of course, only a special case of the 2 x n table dealt with in the preceding section.

A classical example of the application of Chi-square analysis to a 2 x 2 table is provided by the 1913 report of the incidence of typhoid fever in immunized and nonimmunized British military personnel in India (2). The data referred to are presented as the underlined numbers in table 3.

Table 3. Incidence of typhoid fever in immunized and nonimmunized British military personnel. Data from Greenwood and Yule (2).

Personnel	Not attacked	Attacked	Totals
Inoculated (Immunized)	a 10,322 (10,202)	b 56 (176)	10,378
Not inoculated	c 8,664 (8,784)	d 272 (152)	8,936
Totals	18,986	328	19,314

Expected values for each cell of the table (numbers in parentheses) are computed in a manner similar to that previously described.

Cell	Expected value
a	$\frac{10,378 \times 18,986}{19,314} = 10,202$
b	$10,378 - 10,202 = 176$
c	$18,986 - 10,202 = 8,784$
d	$8,936 - 8,784 = 152$

Chi-square then is computed as:

$$\text{Chi-square} = \frac{(10,322 - 10,202)^2}{10,202} + \frac{(56-176)^2}{176} +$$

$$\frac{(8,664-8,784)^2}{8,784} + \frac{(272-152)^2}{152} = 179.61$$

There is only one degree of freedom associated with Chi-square computed from a 2 x 2 table since when an expected value for any cell is computed, the expected value for each other cell is fixed by the first computed value and the marginal totals. The 0.1%

significance level of Chi-square with one degree of freedom is 10.83. Accordingly, the results of Chi-square analysis of this problem is written as

$$\text{Chi-square}_{[1]} = 179.61^{***}$$

The foregoing example also serves to illustrate another method of calculating Chi-square from a 2 x 2 table. The general formula is

$$\text{Chi-square} = \frac{(ad-bc)^2 \times N}{(a+b)(b+d)(c+d)(a+c)}$$

Where N is the grand total and the quantities (a+b) etc., are the marginal subtotals. Substituting the numerical values of the foregoing 2 x 2 table, Chi-square is computed as

$$\text{Chi-square} = \frac{[(10,322 \times 272) - (56 \times 8,664)]^2 \times 19,314}{10,378 \times 328 \times 8,936 \times 18,986} = 180.38.$$

The slight discrepancy between this value of Chi-square and that obtained previously (179.61) resulted from rounding numbers in calculation of expected values in the former case. If enough decimal places had been carried, the values of Chi-square would have been the same.

The values of Chi-square obtained in the three foregoing examples all were unusually large. In fact, careful consideration of the data no doubt would have led to the same conclusions regarding significance even without calculation of Chi-square. If the data had been based on only one-tenth as many individuals in each case, however, the significance of the results would not have been as obvious and the values of Chi-square would have been much less.

The t Test

In a preceding section, the concepts of variance, standard deviation and standard deviation of the mean were discussed briefly. It should be evident that when successive samples of size n are drawn at random from any normally distributed "population" of values of a continuous variable, the means of the several samples will not be equal. In fact these sample means will be normally distributed about the true population mean with standard deviation ($s/\sqrt{n}$). Furthermore, if the mean of a sample differs from the population mean (known or expected on the basis of an hypothesis) by more than about $2s/\sqrt{n}$, we conclude that the sample probably was not truly drawn from the population (i.e., drawn from a different population) or that the hypothesis concerning the population mean was incorrect. The ratio of the difference between the observed sample mean and the expected mean to the standard deviation of the mean is known as Student's t test and is one of the most useful of the tests for statistical significance applicable to data based on measurements, i.e., continuous variables.

Significance of a single mean.--It is indeed a rare occasion when the true mean of a population of values is known but frequently it is desired to determine the significance of a difference between an observed sample mean and some value of the population mean expected in accordance with a certain hypothesis. In many instances the expected value of the mean will be 0, such as in dealing with series of determinations obtained by two methods. If both methods are of equal accuracy, a series of parallel determinations obtained with several samples would provide a population of differences (between values obtained by the two methods) of which the expected mean value is 0.

The general formula for t for use in such problems is

$$t = \frac{\bar{Y} - E}{s/\sqrt{n}}$$

where t is Student's t statistic, $\bar{Y}$ is the observed sample mean, E is the expected value of the mean, and $s/\sqrt{n}$ is the standard deviation of the sample mean.

This application of the t test may be illustrated by a specific example. Davis and Fugo (3) have reported a comparison of two methods for determining the pregnanediol content of urine specimens (Table 4).

Table 4. Pregnanediol content of urine specimens as determined by two methods. Data from Davis and Fugo (3).

Specimen	Pregnanediol content (mg.)	
	Venning method	Modified Guterman method
1	78.1	78.3
2	68.4	60.1
3	65.7	67.9
4	18.2	16.3
5	9.1	8.1

For calculation of t, the data are expressed as a series of differences from which the mean difference is determined.

Method		Difference (d)
V	G	(V-G)
78.1	78.3	-0.2
68.4	60.1	8.3
65.7	67.9	-2.2
18.2	16.3	1.9
9.1	8.1	1.0
Σ d		8.8
Mean difference ($\bar{d}$)		1.76

The standard deviation of the mean difference is calculated in the following manner:

$$\text{Variance } (s^2) = \frac{(-0.2)^2 + (8.3)^2 + (-2.2)^2 + (1.9)^2 + (1.0)^2 - \frac{(8.8)^2}{5}}{4}$$

$$= \frac{78.38 - 15.49}{4} = 15.72$$

$$\text{Std. dev. } (s) = \sqrt{15.72} = 3.97$$

$$\text{Std. dev. of mean } (s_{\bar{d}}) = \frac{3.97}{\sqrt{5}} = \frac{3.97}{2.24} = 1.77$$

t then is calculated as

$$t\ [4] = \frac{1.76 - 0}{1.77} = 0.99$$

In this problem, t was computed from a series of 5 independent values (differences) and one restriction was placed on the series,—they were summed to permit calculation of the mean difference. Accordingly, there are four degrees of freedom associated with this value of t.

Referring to the table of values of t (Table II, appendix A), it is found that, with four degrees of freedom, a value of t as large as 2.78 will occur by chance alone once in 20 trials (5% level). The value of t obtained in this problem, 0.99, would be exceeded by pure chance alone so frequently that it could not possibly be considered significant. Accordingly the null hypothesis, that there is no true difference in the pregnanediol determinations by the two methods, is accepted.

It will be recognized that this use of the t test generally is applicable to data obtained from paired experiments.

Significance of difference between two means.—The t test is more frequently employed in testing the significance of a difference between two means, i.e., between means of two samples or series. The general procedure is similar to the foregoing but the best computational method to follow is dependent upon whether a large (30 or more) or a small (less than 30) number of individual values are included in the sample.

Small samples: The formula for calculation of t with small samples is as follows:

$$t = \frac{\bar{Y}_1 - \bar{Y}_2}{s}\sqrt{\frac{n_1 \times n_2}{n_1 + n_2}}$$

where $\bar{Y}_1$ and $\bar{Y}_2$ are the respective means of the two series, s is the estimate of the pooled standard deviation, and n_1 and n_2 are the number of individuals in the two series.

Data reported by Calder (4) relative to the effect of nicotinic acid on the respiratory rate of rats at a barometric pressure of 300 mm Hg are satisfactory for illustration of the use of t for testing the significance of a difference between two means. The portion of the data selected are presented in Table 5.

Table 5. Respiratory rate of treated (nicotinic acid) and untreated rats at a barometric pressure of 300 mm Hg. Selected from data by Calder (4).

Respiratory rates of Rats	
1. Treated group (1 mg. nicotinic acid per rat)	2. Untreated group (controls)
108	120
126	174
240	192
132	192
108	216
156	186
144	162
132	252
108	150
156	192
Σ 1410	1836
n 10	10
Mean 141.0	183.6

The pooled variance and pooled standard deviation are determined from the individual values as follows:

$$\text{Variance } (s^2) = \frac{\left(108^2 + 126^2 + \cdots 156^2 - \frac{1410^2}{10}\right) + \left(120^2 + 174^2 + \cdots 192^2 - \frac{1836^2}{10}\right)}{10 + 10 - 2}$$

$$= 1421.8$$

$$\text{Std. dev. } (s) = \sqrt{1421.8} = 37.7$$

t then, is calculated as

$$t\,[18] = \frac{183.6 - 141.0}{37.7} \sqrt{\frac{10 \times 10}{10 + 10}}$$

$$= 2.53$$

It will be noted that 18 degrees of freedom (10 + 10 - 2) were used in calculation of the variance. There were 20 original values

but each series was totalled independently in calculation of the pooled variance. Thus, two restrictions were imposed and only 18 degrees of freedom remain.

The table value of t with 18 degrees of freedom at the 5% level is 2.10. Accordingly, the value of 2.53 obtained in the foregoing example exceeds this level of significance; actually it almost equals the value for the 2% level of significance, 2.55. The effect of nicotinic acid, therefore, is considered significant since a difference as great as that observed would be expected to arise by chance alone only about once in 50 trials.

Large samples: The value of t can be calculated from large samples in the same manner as presented in the preceding section. However, as larger and larger numbers of individuals are included in the series, the effect of dividing the sums of squares of deviations from the means by n-1 rather than by n in calculating variances becomes less important and the calculation can be simplified. t can be calculated as

$$t = \frac{\bar{Y}_1 - \bar{Y}_2}{s_d}$$

where $\bar{Y}_1 - \bar{Y}_2$ denotes the difference in the means and s_d is the standard error of the difference. The standard error of the difference is calculated from the variances of the two series of values (designated as 1 and 2) as follows:

$$s_d^2 = \frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}$$

$$\text{then, } s_d = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Space will not be devoted here to full presentation of an example based upon sufficient data for illustration of the computational procedure. However, the data presented in Table 5, and the subsequent calculations, will serve to illustrate the method.

The sum of squares of deviations from the means and the variances (using n as the divisor) are

$$\text{Treated group (1)} = \frac{13,914.0}{10} = 1391.40$$

$$\text{Control group (2)} = \frac{11,678.4}{10} = 1167.84$$

$$\text{then } s_d^2 = \frac{1391.40}{10} + \frac{1167.84}{10} = 255.92$$

$$\text{and } s_d = \sqrt{255.92} = 16$$

$$\text{finally, } t = \frac{42.6}{16} = 2.66$$

This value of t, calculated by the method appropriate for large samples, is very little different from that previously obtained (2.53) even though small sample data were employed.

Confidence limits.--A particularly useful application of the t statistic is in the estimation of confidence limits. These are limits within which it is estimated that the true value of the mean lies, provided the information upon which the mean value is based is reliable. The breadth of these limits will depend, of course, upon the degree of certainty or "confidence" it is deemed necessary to require.

Confidence limits, L, of a single mean are calculated as

$$\pm L = \pm t \left(s/\sqrt{n} \right)$$

where t is the table value of t at any chosen level of probability with the appropriate degrees of freedom, and $s/\sqrt{n}$ is the standard deviation of the mean. The 95% confidence limits of the mean difference calculated from the data of Table 4 thus are

$$\pm L = \pm 2.78 \times 1.77 = \pm 4.92$$

It can be concluded therefore, with a probability of being wrong of only 0.05, that the true mean difference lies between 1.76 ± 4.92, or between -3.16 and 6.68.

Confidence limits for a difference between two means is calculated in a similar manner:

$$\pm L = \pm t \left(s_{\bar{d}} \right)$$

where t has the value determined as before and $s_{\bar{d}}$ is the standard deviation of the mean difference determined as

$$s_{\bar{d}} = s\sqrt{\frac{n_1 + n_2}{n_1 \times n_2}}$$

The 95% confidence limits of the mean difference obtained from the data of Table 5 are determined as

$$\pm L = \pm 2.10 \times 37.7 \sqrt{\frac{10 + 10}{10 \times 10}} = \pm 35.24$$

and it is estimated, with 95% confidence of being correct, that the true mean difference lies between 42.6 ± 35.24, or between 7.36 and 77.84.

The Analysis of Variance

The most widely applicable, and generally most useful technique for the statistical analysis of data is that known as the analysis of variance. This technique, developed by Fisher, in many ways is similar to the t test in that it provides for tests of the significance of differences between means, but it is broader in scope as the significance of differences between means of several groups can be determined concomitantly.

The technique of analysis of variance is possible because of the additive property of variances. That is, if a series of individuals are subjected to variation from multiple independent sources, the total variance is the sum of the individual variances from each source. Accordingly, with appropriate analytical procedures, the total variation in the entire sequence can be broken down or segregated into parts; segments of the variation can be identified with their sources; and, the significance of each component part can be assessed independently.

Between group and within group variation.— The general approach to analysis of data by the analysis of variance involves segregation of the total variation into that resulting from differences between groups and from differences within groups. From these segregated portions of the variation associated with specific groups or with individuals within groups, separate estimates of the population variance can be obtained. If the total series of individuals is merely a random sample of a normally distributed homogeneous population, each estimate of variance should be essentially the same, any actual differences resulting from sampling. If the estimates of variance are significantly different, however, e.g., if the population variance estimated from differences between groups is significantly greater than that estimated from differences between individuals within groups, it is concluded that the entire series is not homogeneous but that true differences in groups actually do exist.

The variance ratio (F test).— The test of significance employed in the analysis of variance is the variance ratio, symbolized by F, and computed as the ratio of two estimates of variance (mean squares). Each estimate of variance or mean square is obtained by dividing the appropriate segment of the total sum of squares of deviations from the mean by the corresponding degrees of freedom.

Analysis of variance of single classification data.— Single classification data, as used here, refers to that obtained from experiments in which individuals, in groups, have been subjected to a variety of treatments (or levels) while all other known factors considered of importance have been held constant. The significance of differences between the average responses of any two groups, of course, can be tested by the t test presented in the preceding section, but the significance of differences between all groups can be determined at the same time by analysis of variance.

Data relative to the weight of thyroids of mice, obtained by Keating and Soberman (5) while studying the uptake of radioiodine by mice made obese by injection of goldthioglucose, are selected for illustration of this type of statistical analysis. Individual thyroid weights are given for three groups of mice: (A) treated, obese; (B) treated, nonobese; and (C) controls (Table 6).

Table 6. Thyroid weights of normal mice and of obese and nonobese mice resulting from injection of goldthioglucose. Data from Keating and Soberman (5).

Thyroid weights (mg.)		
A. Treated, obese	B. Treated, nonobese	C. Controls
5.5	1.8	3.8
2.5	1.5	2.2
6.3	2.1	2.1
7.0	3.5	1.3
5.6	2.4	1.5
4.1	2.1	1.7
2.9	1.9	1.6
2.5	2.2	1.9
2.1	1.8	
4.5	1.2	
Σ 43.0	20.5	16.1
n 10	10	8
Means 4.3	2.5	2.01

Successive steps in the computation of quantities required for the analysis of variance are

Grand total (G) = $43.0 + 20.5 + 16.1 = 79.6$

$N = n_A + n_B + n_C = 10 + 10 + 8 = 28$

Correction factor (C.F.) = $\frac{G^2}{N} = \frac{79.6^2}{28} = 226.29$

Total sum of squares (S.S.) = $5.5^2 + 2.5^2 + \cdots + 1.9^2 - C.F.$

$= 294.82 - 226.29 = 68.53$

Between groups S.S.[4] $= \frac{43.0^2}{10} + \frac{20.5^2}{10} + \frac{16.1^2}{8} - C.F.$

$= 259.32 - 226.29$

$= 33.03$

Within groups S.S. = Total S.S. - Between groups S.S.

$= 68.53 - 33.03 = 35.50$

---

[4] In calculating the sum of squares between groups, the square of the total of each group is divided by the number of individuals in each group.

The computed values are entered in an analysis of variance table of the following form:

Analysis of variance (ANOV)

Source of variation (VS)	Degrees of freedom (D.F.)	Sums of squares (S.S.)	Mean squares (M.S.)	Variance ratio (F)
Between groups	2	33.03	16.52	11.63***
Within groups (Error)	25	35.50	1.42	
Total	27	68.53		

In the above analysis of variance table, it will be seen that the two sources of variation, between and within groups (experimental error) have been segregated. The total number of degrees of freedom (second column) is 27, one less than the original number of observations since a grand total or summation was required for calculation of the correction factor. Two of these degrees of freedom are associated with the variation between groups (3 groups -1 = 2) and 25 degrees of freedom are associated with variation within groups (A, 10-1 = 9; B, 10-1 = 9; C, 8-1 = 7: $9 + 9 + 7 = 25$).

The sums of squares (third column) were obtained in the manner previously presented in detail. Actually, in this case, the within groups sum of squares (of deviations from the mean) was obtained by subtraction but it can be obtained directly, as follows:

Group A: $S.S. = 5.5^2 + 2.5^2 + \cdots + 4.5^2 - \frac{43.0^2}{10} = 27.78$

Group B: $S.S. = 1.8^2 + 1.5^2 + \cdots + 1.2^2 - \frac{20.5^2}{10} = 3.43$

Group C: $S.S. = 3.8^2 + 2.2^2 + \cdots + 1.9^2 - \frac{16.1^2}{8} = 4.29$

The sum of these three sums of squares ($27.78 + 3.43 + 4.29$) = 35.50, the same value as obtained previously by subtraction.

The mean squares (fourth column) are obtained by dividing the sums of squares by the corresponding degrees of freedom. These are estimates of the population variance, derived from groups (means) and from individuals within groups. The two estimates of variance are markedly different, so different that it is readily apparent that it is highly improbable that they are estimates of the same population variance.

The variance ratio (F), in the last column, is obtained by dividing the larger mean square (between groups, 16.52) by the smaller mean square

(within groups or error, 1.42). The values of F required for significance are tabulated in Table III (appendix A). In using this table, it is necessary to take into consideration the degrees of freedom associated with each mean square. The ratio always is computed with n_1 degrees of freedom (across the top of the table) denoting degrees of freedom of the larger mean square and n_2 degrees of freedom (left hand column of the table) denoting those of the lesser mean square. Entering the table with 2 (n_1) and 25 (n_2) degrees of freedom, respectively, it is found that a value of F as large as 9.2 will arise by chance alone only once in 1000 times if there is no true difference between the treatment groups (null hypothesis). The value of F obtained in this problem, 11.63, clearly is highly significant.

Analysis of variance of multiple classification data.— In many experiments, more complex than that just presented, the study subjects as a whole deliberately are subjected to more than one type of controlled variable or experimental condition, or the study subjects themselves are varied or selected, e.g., segregation of experimental animals by sex, or into weight classes, etc. Regardless of the particular nature of the experiment, the individuals (observations) are subject to variation from multiple sources. Provided the design of the experiment is appropriate, it often is possible to segregate those portions of the total variation associated with each factor (or level) and, furthermore, to determine the dependency of the effectiveness of one factor on another (interaction).

Data for illustration of the application of analysis of variance to a more complex experiment have been selected from a study reported by Stefanini (6) on the effect of three chemical agents on the sedimentation rates of human and dog bloods (Table 7).

Table 7. Sedimentation rates (in mm) of human and dog bloods treated with citrate, heparin and Amberlite. Data selected from findings reported by Stefanini (6).

Chemical agent	Source of blood		Totals
	Human	Dog	
Citrate	10.0 14.5 23.0 17.5	2.0 2.0 2.5 3.0	74.5
Heparin	9.7 15.0 24.0 18.0	2.0 2.0 3.0 3.0	76.7
Amberlite	4.0 9.0 15.0 10.5	1.0 1.0 1.0 2.0	43.5
Totals	170.2	24.5	194.7

Preparatory to performing an analysis of variance on data from a complex experiment, it is well to determine the degrees of freedom and the portions of the total variation with which they are associated. In this problem, there are 24 separate entries or observed values. Accordingly, there must be a total of 23 degrees of freedom. Each of the six cells of the table include four independent values; each cell then contributes three degrees of freedom, (4-1 = 3)—a total of 18. There are two sources of blood, human and dog, which accounts for one degree of freedom (2-1 = 1). Differences between the three chemical agents contribute two degrees of freedom (3-1 = 2). The remaining two degrees of freedom are associated with the interaction between chemical agents and sources of blood, i.e., with failure of differences in the effects of the chemical agents to be comparable with the bloods from the two sources. Degrees of freedom associated with an interaction always equal the product of the degrees of freedom of the main effects involved (in this case, 2 x 1 = 2).

The computational procedure is similar to that presented in the previous example. The first steps are as follows:

Correction factor $= \frac{194.7^2}{24} = 1,579.50$

Total S.S. $= 10.0^2 + 14.5^2 + \cdots\cdots 2.0^2 - C.F.$

$= 2,853.09 - 1,579.50$

$= 1,273.59$

Between bloods S. S. $= \frac{170.2^2}{12} + \frac{24.5^2}{12} - C.F.$

$= 2,464.02 - 1,579.50$

$= 884.52$

Between chemicals S.S. $= \frac{74.5^2}{8} + \frac{76.7^2}{8} + \frac{43.5^2}{8} - C.F.$

$= 1,665.67 - 1,579.50$

$= 86.17$

For calculation of the sum of squares associated with the interaction, a new table is constructed in which only the cell totals (sum of the 4 individuals) are entered. These cell totals commonly are referred to as subclass totals.

Chemical agent	Source of blood		Totals
	Human	Dog	
Citrate	65.0	9.5	74.5
Heparin	66.7	10.0	76.7
Amberlite	38.5	5.0	43.5
Totals	170.2	24.5	194.7

A subclass sum of squares then is calculated as

$$\text{Subclass S.S.} = \frac{65.0^2 + 66.7^2 + \cdots + 5.0^2}{4} - \text{C.F.}$$

$$= 2{,}592.85 - 1{,}579.50$$

$$= 1{,}013.35$$

This subclass sum of squares differs from the total sum of squares in that differences between individuals within subclasses have been eliminated by the summation. Accordingly, the subclass sum of squares includes the sums of squares between blood sources, between chemical agents, and that associated with the interaction. Since the first two have been determined, the latter can be obtained by subtraction.

$$\text{Interaction S.S.} = \text{Subclass S.S.} - (\text{Bloods S.S.} + \text{Chemicals S.S.})$$

$$= 1{,}013.35 - (884.52 + 86.17)$$

$$= 42.66$$

Calculation of the within groups or error sum of squares can be performed directly by adding the values obtained from each of the six cells of the table in the following manner:

$$\text{First cell S.S.} = 10.0^2 + 14.5^2 + 23.0^2 + 17.5^2 - \frac{65.0^2}{4}, \text{ etc., but it is}$$

easier to obtain the sum of these values by direct subtraction of the known sums of squares from the total sum of squares.

The computed values then are entered into an analysis of variance table as shown below.

ANOV

V. S.	D. F.	S. S.	M. S.	F.
Between bloods	1	884.52	884.52	61.17***
Between chemicals	2	86.17	43.08	2.98
Bloods x chemicals (Interaction)	2	42.66	21.33	1.48
Within subclasses (Error)	18	260.24	14.46	
Total	23	1,273.59		

Interpretation of an analysis of variance often is difficult. In this particular case, the difference between the bloods (human and dog) clearly is

significant. Differences between the chemicals fail to prove significant since the 5% significance level value of F with 2 and 18 degrees of freedom is 3.6. Yet the value of F obtained, 2.98, is sufficiently large to suggest that a difference actually does exist. Failure to attain significance in this case resulted in part from marked differences between individuals within subclasses and, also, to some extent, from inconsistency in the effects of the chemicals with the two types of blood employed. This inconsistency or discrepance is the same as the interaction which was segregated and assessed. The ratio of the variances for interaction and error likewise failed to prove significant.

There is no unanimity of opinion as to the proper course of action when an interaction is found to be of questionable significance. If the interaction had been significant it would have become in fact the error term for testing the significance of the main effects. When of questionable significance, some investigators would pool its sum of squares with the error sum of squares and obtain a new estimate of error variance based on 20 degrees of freedom. Others would be inclined to split the experiment into two parts and to consider the effects of the chemicals on each type of blood separately.

The foregoing problem was selected for presentation for the main purpose of illustrating methods of computation and analysis of an experiment of this degree of complexity. The final analysis of variance, however, is somewhat misleading and extension of the analysis is considered advisable. The segregation of the total sum of squares into portions associated with differences between chemicals, and differences within subclasses, for example, is strictly adequate only when interest is directed equally to all three chemical agents. On this basis, the differences between chemicals did not prove significant.

In the particular study from which these data were selected, the investigator specifically was interested in the chemical Amberlite and in its effect as compared with that of citrate and heparin. Accordingly, it is not necessary to direct consideration to all three chemicals in the sense that they represent a sample of a population of chemicals. When that is done, our attention is directed to the estimate of population variance based on two degrees of freedom.

A specific segment of the total sum of squares is associated with each degree of freedom and, in a properly designed experiment, a free and independent comparison (orthogonal comparison) can be made for each degree of freedom. Thus, in the case at hand, since there are two degrees of freedom between chemicals, two comparisons between chemicals can be made independently of each other. The first comparison of interest is that between Amberlite and the average of the other two chemicals, heparin and citrate. With this comparison selected there is only one other independent comparison that can be made, that is a comparison of heparin and citrate.

Sums of squares corresponding to the two individual degrees of freedom associated with the selected comparisons are calculated as follows:

Amberlite vs (heparin and citrate)—

$$S.S. = \frac{43.5^2}{8} + \frac{(76.7 + 74.5)^2}{16} - \frac{194.7^2}{24} = 85.87$$

Heparin vs citrate —

$$S.S. = \frac{76.7^2}{8} + \frac{74.5^2}{8} - \frac{151.2^2}{16} = 0.30$$

These two sums of squares total 86.17 which is identical with that previously obtained when the two degrees of freedom were not separated.

The analysis of variance table now can be reconstructed substituting these two sums of squares for the "between chemicals" sum of squares previously employed.

From the completed analysis, which follows, it is concluded that the effect of Amberlite on the blood sedimentation rate is significantly different than is that of heparin or citrate.

ANOV

V. S.	D. F.	S. S.	M. S.	F
Between bloods	1	884.52	884.52	61.17***
Between chemicals: Amberlite vs. (heparin and citrate)	1	85.87	85.87	5.94 *
Heparin vs. citrate	1	0.30	0.30	—
Interaction	2	42.66	21.33	1.48
Error	18	260.24	14.46	
Total	23	1,273.59		

Independent comparisons.— In the immediately preceding paragraphs two concepts were introduced: (a) that of an association of each individual degree of freedom with a particular segment of the total sum of squares, and (b) that of an association of each degree of freedom with an independent (orthogonal) comparison possible from the data. These associations obviously are important, not only in relation to designing experiments so as to provide for the particular comparison(s) of interest to the investigator, but also in relation to the computational procedure involved in analysis. It is the latter aspect that is dealt with here.

Much of statistical analysis is concerned with assessment of the significance of differences or, in other words, with determination of the certainty of true existence of observed apparent effect as indicated by comparisons or contrasts of the attributes (measurements or enumerations) of groups of individuals under specified experimental conditions. When the mean value of the attributes of such groups are expressed in appropriate numerical units, comparison of the groups (determination of difference) is accomplished by the simple arithmetic expedient of subtraction.

Consider two groups A and B, each comprised of the same number of individuals (k), and their total equal to A + B. Comparison of these groups (determination of difference) is accomplished by simple subtraction, A - B which, of course, is equivalent to (+ 1) A + (-1)B. Specification of the comparison therefore required only the assignment of the coefficients (+ 1) and (-1).

When a third group C is added to the series so that the total equals A + B + C, identification of comparisons is more difficult. Three comparisons between single groups are obvious, viz., A - B, A - C and B - C. However, these comparisons are not independent since A - B numerically equals (A - C) - (B - C). There also are three possible complex comparisons, viz., (A + B) -2C, (A + C) -2B, and 2A - (B + C). It will be found that the first of these, (A + B) -2C, is the only one of these complex comparisons independent of A - B. That is, if the comparison A - B were to take a different value, but the total A + B + C remained constant, the only comparison which would not necessarily be changed is (A + B) -2C. Therefore, these comparisons are independent.

Determination of which comparisons are independent (orthogonal) is not as difficult or laborious as might appear from the foregoing but is accomplished quite easily through employment of coefficients provided the groups are made up of an equal number of individuals (k). Assignment of coefficients is illustrated by the table which follows.

The rules for assuring independence of comparisons are: the sum of the coefficients in each comparison must be 0, and the sum of the cross products of the coefficients in the comparisons also must be 0.

	Experimental groups			
Comparison sets	A	B	C	Sums
Comparison 1	+ 1	- 1	0	0
Comparison 2	+ 1	+ 1	-2	0
Cross product (1 x 2)	+ 1	- 1	0	0
or				
Comparison 1	+ 1	0	- 1	0
Comparison 2	+ 1	- 2	+ 1	0
Cross products (1 x 2)	+ 1	0	- 1	0
etc.				

Calculation of mean squares for two groups.— Incorporation of the use of coefficients in the computational procedures of the analysis of variance requires the introduction of a new method of computing mean squares for single degrees of freedom(difference between two groups). The formula may be written as

$$\text{Mean square} = \frac{(\Sigma Y_1 - \Sigma Y_2)^2}{D}$$

where ΣY_1 and ΣY_2 are totals of groups made up of the same number of individuals (k), and D, the divisor, is the product of the sums of the squares of the coefficients and k.

The computational procedure using coefficients and the new formula for mean squares is illustrated by recalculation of the mean squares from the data of Table 7 (page 32). The data are rearranged to facilitate computation as shown in the chart which follows.

In the computation using coefficients, it will be noted that the mean squares are identical with those obtained by the longer method previously presented, (c.f. ANOV, p.34). The sum of squares for interaction obtained with two degrees of freedom, 42.66, is equal to the sum of the interaction mean squares 0.09 + 42.57.

Calculation of Mean Squares from Data of Table 7

Bloods	Human			Dog			Calculation of mean squares				
Chemicals	C	H	A	C	H	A					
Group Sums	65.0	66.7	38.5	9.5	10.0	5.0	$\Sigma +$	$\Sigma -$	T^2 *	D#	M. S.+
Bloods	+	+	+	-	-	-	170.2	24.5	21,228.49	24	884.52
(C + H) - 2A	+	+	-2	+	+	-2	151.2	87.0	4,121.64	48	85.87
C - H	+	-	0	+	-	0	74.5	76.7	4.84	16	0.30
Bloods x [(C + H) - 2A]	+	+	-2	-	-	+ 2	141.7	96.5	2,043.04	48	42.57
Bloods x (C - H)	+	-	0	-	+	0	75.0	76.2	1.44	16	0.09

* T^2 is identical with the expression $(\Sigma Y_1 - \Sigma Y_2)^2$ in the general formula (page 38).

D is the product of the sums of squares of the coefficients x k.

k, the number of individuals per group, in this case is 4.

Example: D for the comparison (C + H) - 2A = $(+1)^2 + (+1)^2 + (-2)^2 + (+1)^2 + (+1)^2 + (-2)^2 \times 4 = 48$.

+ M. S. = T^2/D.

In this particular problem, the error sum of squares must be calculated as before. Nevertheless, the saving in labor of computation is considerable although this is only one of the advantages of individual degree of freedom analysis. Unfortunately, many experiments are not balanced and the foregoing method of analysis is not feasible.

Correlation

Correlation refers to the relationship between two variables but not so much in the sense of mutual dependency as in the sense of covariation, i.e., the degree to which two variables vary together. Correlation analysis was used in the examination of data far more extensively in the past than it is at present. This loss in popularity has resulted from changes in experimental methods and techniques rather than from any particular inadequacy or unreliability of the statistic.

Correlation analysis is most readily applicable to observational data obtained in situations where possible causative factors may be discerned but not controlled, and attributes appear to vary in occurrence or degree in association with such factors. In present day experimentation, particularly in laboratory investigations where the researcher exercises control over causative factors, other statistical procedures usually are more useful.

The correlation coefficient.— The correlation coefficient r is a convenient statistic used in determining the degree of reliability of an association of two factors. The limiting values of r are + 1 and - 1. Positive values indicate that a change in one variable is associated with a change in the other in the same direction (increase or decrease) while negative values indicate that the associated variations are in opposite direction.

If one variate is denoted by values of X and the other by values of Y, the general formula for the CORRELATION coefficient is

$$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \cdot \Sigma y^2}}$$

where r is the correlation coefficient, Σxy is the sum of the cross-products of the deviations of the individual values of X and Y from their respective means, and Σx^2 and Σy^2 are the sums of squares of these deviates.

Calculation and assessment of significance of the correlation coefficient is illustrated with data reported by Godfrey (7) on the typhoid case fatality rate of patients of different ages (Table 8).

Table 8. Age group and case fatality rates of typhoid fever cases. Data abridged from Godfrey (7).

Age group	Case fatality (percentage)
(X)	(Y)
5 - 9 (1)	4.4
10 - 14 (2)	6.4
15 - 19 (3)	10.2
20 - 24 (4)	12.5
25 - 29 (5)	13.5
30 - 34 (6)	16.4
35 - 39 (7)	15.2
40 - 44 (8)	15.2
45 - 49 (9)	19.9
50 - 54(10)	22.9
Σ 55	Σ 136.6

The numbers 1 to 10 in parentheses have been entered as working values of X to facilitate calculations. In correlation, interest is directed toward the uniformity of the step by step covariation of the two variables, rather than toward the degree of change. The change from midpoint to midpoint of each age interval 5 to 9, 10 to 14, 15 to 19, etc., is constant and if these actual midpoints were used instead of the numbers 1, 2, 3, etc., the numerical value of r would be the same.

The necessary calculations are performed as follows:

$$\Sigma xy = \Sigma XY - \frac{(\Sigma X)\ (\Sigma Y)}{n}$$

$$= (1 \times 4.4) + (2 \times 6.4) + \cdots (10 \times 22.9) - \frac{55 \times 136.6}{10}$$

$$= 899.8 - 751.3$$

$$= 148.5$$

$$\Sigma x^2 = 1^2 + 2^2 + 3^2 + \cdots 10^2 - \frac{55^2}{10}$$

$$= 385 - 302.5$$

$$= 82.5$$

$$\Sigma y^2 = 4.4^2 + 6.4^2 + 10.2^2 + \text{---} + 22.9^2 - \frac{136.6^2}{10}$$

$$= 2154.32 - 1865.96$$

$$= 288.36$$

then

$$r = \frac{148.5}{\sqrt{82.5 \times 288.36}}$$

$$= \frac{148.5}{154.2}$$

$$= 0.96$$

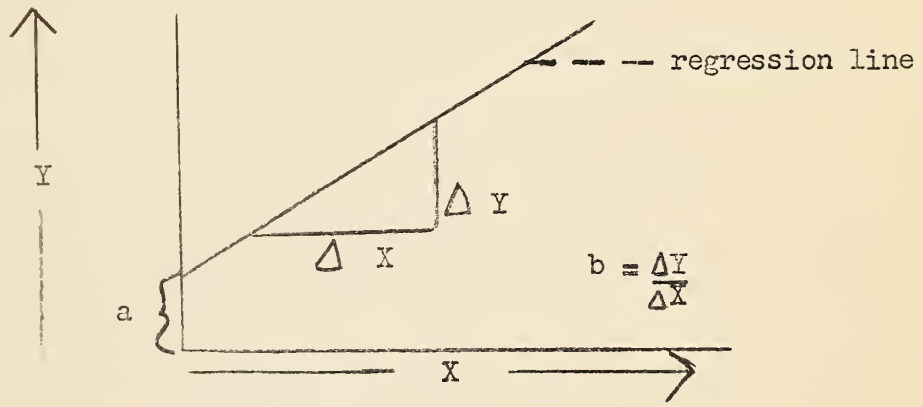
There are n - 2 degrees of freedom associated with r. n designates the number of pairs, 10 in this problem, and two restrictions have been imposed since r really is a test of the linearity of the relationship of the form Y = a + b X where a and b are constants (restrictions). By reference to Table IV (appendix A) it is seen that a value of r as great as 0.96 would arise by pure chance alone less frequently than once in 1000 times.

Linear Regression

Regression also refers to the relationship of two variables but differs from correlation in that it deals with the degree of change in one (dependent variable, Y) associated with or occasioned by a certain change in the other (independent variable, X). When this quantitative relationship is expressed algebraically, it is possible to predict values of the dependent variable at any given value of the independent variable.

Regression may be linear (straight line relationship between X and Y) or curvilinear (curved line relationship). Many biological relationships are curvilinear when the variables are expressed in arithmetic units but are linear (at least approximately so) when expressed in other units such as logarithms, probits, etc. Conversion of arithmetic data to other units for such purpose is referred to as transformation of the data.

In consideration of linear regression it is helpful to think of a straight line on a graph passing through or among points representing values of X (horizontal axis) and associated values of Y (vertical axis).



The equation of a straight line is

$$Y = a + b\,X$$

where a is the intercept of the Y axis and b (the ratio of change in Y to change in X) is the slope of the line.

The regression line is that line which is so located that the sum of the squares of the deviations of the observed from the expected values of Y is minimal (assuming X is measured without error). This "least squares" regression line is best calculated in the following form

$$\hat{Y} = \bar{Y} + b\,(X - \bar{X})$$

where $\hat{Y}$ is the expected value of Y at a given value of X, $\bar{Y}$ is the mean of the observed values of Y, b is the regression coefficient or slope of the line computed as $\frac{\Sigma xy}{\Sigma x^2}$, and $\bar{X}$ is the mean of the given values of X.

The computational procedure is illustrated with data by Brinkhaus (8) on the effect of centrifugation time on the clotting time of recalcified normal plasma (Table 9).

Table 9. Effect of centrifugation time on clotting time of recalcified normal plasma. Data selected from Brinkhaus (8).

Centrifugation time (minutes)	Clotting time (minutes)
(X)	(Y)
10	8.5
30	10.0
60	18.5
90	20.0
120	31.0
165	37.0

The quantities required for calculation of the regression equation are computed as follows:

$$\bar{Y} = \frac{\Sigma Y}{n} = \frac{125}{6} = 20.83$$

$$\bar{X} = \frac{\Sigma X}{n} = \frac{475}{6} = 79.17$$

$$\Sigma xy = \Sigma XY - \frac{(\Sigma X)\ (\Sigma Y)}{n}$$

$$= 13{,}120.00 - 9{,}895.83$$

$$= 3{,}224.17$$

$$\Sigma x^2 = \Sigma X^2 - \frac{(\Sigma X)^2}{n}$$

$$= 54{,}325.00 - 37{,}604.17$$

$$= 16{,}720.83$$

$$b = \frac{\Sigma xy}{\Sigma x^2} = \frac{3{,}224.17}{16{,}720.83} = 0.19$$

The equation then is calculated as

$$\hat{Y} = 20.83 + 0.19\ (X - 79.17)$$

$$= 20.83 - 15.04 + 0.19X$$

$$= 5.79 + 0.19X$$

Then by substituting any appropriate value of X (centrifugation time) into the equation, $\hat{Y} = 5.79 + 0.19X$, the expected corresponding value of Y (clotting time) can be calculated. However, it usually is not wise to attempt to predict values outside the range covered by the data from which the equation is calculated.

The deviation of the observed values of Y from those expected by the equation can be determined by direct calculation. For example; the expected value of Y at X = 10 is found as follows

$$\hat{Y} = 5.79 + 0.19(10)$$

$$= 7.69$$

The difference between the observed and expected values of Y at X = 10 then is 8.5 - 7.69 = 0.81. In this problem, all observed and expected values of Y are in good agreement as shown below

X	Observed Y Y_o	Expected Y $\hat{Y}$	$Y_o - \hat{Y}$	$(Y_o - \hat{Y})^2$
10	8.50	7.69	0.81	0.66
30	10.00	11.49	-1.49	2.22
60	18.50	17.19	1.31	1.72
90	20.00	22.89	-2.89	8.35
120	31.00	28.59	2.41	5.81
165	37.00	37.14	-0.14	0.02
			Total	18.78

If any other value of b had been used in calculation of the expected values of Y, the quantity $(Y_o - \hat{Y})^2$ would have been greater than 18.78 (allowing for minor errors resulting from rounding numbers, of course).

Linear regression and analysis of variance.— The general significance of a linear regression or, in other words, the adequacy with which the relationship is described by the linear equation, is readily determined by analysis of variance.

Employing the data of the previous problem, the total sum of squares is

$$\text{Total S. S.} = 8.5^2 + 10.0^2 + \cdots + 37.0^2 - \frac{125^2}{6}$$

$$= 3{,}244.50 - 2{,}604.17$$

$$= 640.33$$

The sum of squares accounted for by the regression is

$$\text{Regression S. S.} = \frac{(\Sigma xy)^2}{\Sigma x^2} = \frac{3{,}224.17^2}{16{,}720.83} = 621.69$$

An analysis of variance table is constructed in the usual manner

ANOV

V. S.	D. F.	S. S.	M. S.	F
Regression	1	621.69	621.69	133.41***
Deviations from regression	4	18.64	4.66	
Total	5	640.33		

In the above analysis of variance, the degrees of freedom and the sum of squares for deviations from regression both are obtained by subtraction. Note that the sum of squares for deviations from regression numerically is the same as the quantity $(Y_o - \hat{Y})^2$ previously obtained (the minor discrepancy resulted from rounding numbers).

The four degrees of freedom for deviations from regression actually represent higher order or curvilinear regressions (quadratic, cubic, etc.). These can be calculated directly but it is a laborious task. If the values of X had been evenly spaced, sums of squares for the individual higher regressions could have been calculated with ease by employing coefficients. This method will be illustrated later with suitable data.

The highly significant value of F obtained in this analysis can be interpreted as meaning that if the relationship between X and Y had not been essentially linear, the chance of obtaining so large a value of F would have been less than 1 in 1000.

Regression, transformation, and analysis of variance with coefficients.— It was stated in preceding sections that when successive values of the independent variable are equally spaced, analysis of variance of the data could be accomplished with the use of coefficients; also that certain non-linear relationships could be made linear when the data are expressed in appropriate units (transformed).

Examples of the foregoing are afforded by data by Tanner et al., (9) on the relation of the initial glucose supply to the synthesis of riboflavin by Ashbya gossypii (Table 10).

Table 10. Initial glucose supply and riboflavin synthesis by A. gossypii. Data selected from Tanner et al., (9).

Initial glucose (%)	Riboflavin (^{ug}/ml in 6 days)
(X)	(Y)
0.25	10.5
0.50	23.0
1.00	35.0
2.00	159.0
4.00	339.0

On preliminary examination of the data it is apparent that there is a relationship between the two variables but that the relationship is not linear. However, it appears that as the initial glucose level is doubled in successive levels, the yield of riboflavin is approximately doubled. This suggests that the two may be linearily related if the data pertaining to each variable are transformed to logarithms. To simplify the computational procedure the values of X (glucose supply) are first multiplied by 10 so that none of the logarithms will be negative. The data, in logarithms, then are as follows

log-(initial glucose x 10)	log-riboflavin (^{ug}/ml)
(X)	(Y)
0.398	1.021
0.699	1.362
1.000	1.544
1.301	2.201
1.602	2.530
$\Sigma X = 5.000$	$\Sigma Y = 8.658$

The regression equation is calculated in the manner previously presented. The necessary computed values are

$$\bar{Y} = \frac{8.658}{5} = 1.732$$

$$\bar{X} = \frac{5}{5} = 1$$

$$\Sigma x^2 = 5.906 - 5.000 = 0.906$$

$$\Sigma xy = 9.819 - 8.658 = 1.161$$

$$b = \frac{\Sigma xy}{\Sigma x^2} = \frac{1.161}{0.906} = 1.281$$

and

$$\hat{Y} = 1.732 + 1.281 (X - 1)$$

$$= 0.451 + 1.281X$$

The expected yield of riboflavin at any initial glucose concentration, say 2%, can be calculated as follows:

$$X = \log-(2 \times 10) = 1.301$$

$$\hat{Y} = 0.451 + 1.281 \times 1.301$$

$$= 2.118$$

The expected yield of riboflavin (in $^{ug}/ml$) at this level of glucose is the anti-log of 2.118 or 131.3 as compared with the observed value of 159.0.

The transformed data can be examined by analysis of variance in the usual manner as follows:

$$\text{Total S. S.} = \Sigma y^2 = \Sigma Y^2 - \frac{(\Sigma Y)^2}{n} = 1.535$$

$$\text{Regression S. S.} = \frac{(\Sigma xy)^2}{\Sigma x^2} = \frac{1.348}{0.906} = 1.488$$

ANOV

V. S.	D. F.	S. S.	M. S.	F
Regression Dev. from regression	1 3	1.488 0.047	1.488 0.016	93***
Total	4	1.535		

It is apparent, therefore, that the relationship of log- (initial glucose x 10) to log-riboflavin synthesis is linear.

The analysis of variance can be accomplished more easily and with a greater yield of information by using coefficients. Tables of coefficients and divisors for sets of orthogonal comparisons in regression are available, the following being taken from Snedecor (10).

Number of levels (X)	Degree of polynomial									
	1	1	2	1	2	3	1	2	3	4
1	-	-	∔	-3	∔	-	-2	∔ 2	-	∔
2	∔	0	-2	-	-	∔3	-	-	∔2	-4
3		∔	∔	∔	-	-3	0	-2	0	∔6
4				∔3	∔	∔	∔	-	-2	-4
5							∔2	∔2	∔	∔
Divisors	2	2	6	20	4	20	10	14	10	70

The transformed data from Table 10 and coefficients are entered into a table and each regression component (degree of freedom) is calculated as follows:

Regression	Values of Y at levels of X					$\Sigma +$	$\Sigma -$
	1.021	1.362	1.544	2.201	2.530		
Linear	-2	-1	0	∔1	∔2	7.261	3.404
Quadratic	∔2	-1	-2	-1	∔2	7.102	6.651
Cubic	-1	∔2	0	-2	∔1	5.254	5.423
Quartic	∔1	-4	∔6	-4	∔1	12.815	14.252

The mean squares for each level of regression are calculated from the columns $\Sigma +$ and $\Sigma -$ in this manner

Regression	$[(\Sigma +) - (\Sigma -)]^2$	Divisor	Mean Square
Linear	14.876	10	1.488
Quadratic	0.203	14	0.014
Cubic	0.029	10	0.003
Quartic	2.065	70	0.030

It will be noted that the mean square for linear regression 1.488 is the same as that previously obtained. The sum of the mean squares for higher order

regressions (0.014 + 0.003 + 0.030 = 0.047) equals the sum of squares previously obtained for deviations from regression (C. F., p.49). When this total is divided by three (degrees of freedom) the pooled error term of 0.016 is obtained. Thus, the two methods yield the same results but by use of the coefficients it is possible to gain information on each of the regressions and the total effort in computation was less. Unfortunately, much data useful for regression analysis are not determined employing successive values of X differing by constant increments. Even in the problem presented here it was necessary to omit part of the original data to permit use of the method of coefficients.

The Poisson Distribution

The Poisson distribution previously was identified (page 15) as that type in which the variance equals the mean. While the Poisson distribution is of interest primarily in consideration of events of rare occurrence, certain types of data are not infrequently obtained in research in which this relationship apparently exists. It is well to be familiar with the distribution at least to an extent sufficient to permit its identification.

In bacteriology and immunology it frequently is desired to administer pathogenic organisms to experimental animals in doses just sufficient to kill all or a given proportion of a group. The most commonly employed dose is that which is estimated to kill 50% of the animals, the L. $D._{50}$ (lethal dose for 50%). When highly virulent organisms are employed, such doses may be quite small in terms of actual number of organisms. In such cases, experimental findings often are poorly reproducible. The source of at least part of this poor reproducibility can be identified through study of the distribution of small numbers.

In a study of the degree of variation in small doses of Salmonella typhosa used in typhoid vaccine potency tests, the author (11) prepared a total of 176 pour plates from a dilution estimated to contain five viable organisms per ml. The viable counts observed following appropriate incubation are given in Table 11.

The expected frequency of each count, in accordance with the Poisson distribution, is given in the third column of the table.

Table 11. Observed and expected frequency of viable S. typhosa as determined by plate counts. Data from Batson (11).

Viable count (Colonies per plate)	Observed frequency (No. of plates)	Expected frequency	Chi-square
0	2	n/e^m = 2.06	)
1	9	$mn/e^m(1!)$ = 9.04	) 0.0009
2	17	$m^2n/e^m(2!)$ = 20.36	0.5545
3	30	$m^3n/e^m(3!)$ = 30.20	0.0013
4	29	$m^4n/e^m(4!)$ = 33.60	0.6298
5	44	$m^5n/e^m(5!)$ = 29.90	6.6492
6	20	$m^6n/e^m(6!)$ = 22.18	0.2143
7	11	$m^7n/e^m(7!)$ = 14.10	0.6816
8	11	$m^8n/e^m(8!)$ = 7.84	)
9	1	$m^9n/e^m(9!)$ = 3.88	) 0.0225
10	2	$m^{10}n/e^m(10!)$ = 1.73	)
Totals	176	174.89	8.7541#

n (number of plates) = 176
m (average count) = 4.45
Table value of Chi-square 0.05 (6 df) = 12.59

Calculation of these values from the successive elements of the series appears quite formidable but actually is not at all difficult. The calculation involves mainly successive additions and subtraction of logarithms. Required values are:

n = number of plates = 176. log n = 2.2455

e = base of natural logarithms = 2.71828. log e = 0.4343

m = average count = 4.45. log m = 0.6484

and logarithms of the numbers, 1, 2, ——— 10.

M. Potter

[illegible]

[illegible]

[illegible]

[illegible]

Table 1

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

Table 2

Typhoid Mouse Protection Potency Assay of

Typhoid Vaccine

(survivors/total mice at 72 hours)

Volume and doses in ml.				
Unknown		Reference		Zone
0.0015	0.15	0.015	0.15	[illegible]
1/20	16/20	1/20	17/20	[illegible]

Assay procedure: Immunizing doses, I.P.
Challenge after 5 days with 5000 [illegible]
[illegible] in 5% mucin, I.P.

Table 2

Chi-square analysis of 2-dose Typhoid Vaccine Potency Test

Degree of freedom	Vaccines, dosage levels and survivors per 20 mice				Calculation of chi-square*						
	Unknown		Reference								
	D2	D2	D1	D0							
	4	18	1	17	Σa	$\Sigma -$	T	T^2	D	T^2/D	Level of significance
Test of fit	5	4	-	-	22	18	4	16	80	0.20	0.65
Dose effect	0	-	4	-	5	35	30	900	80	11.25	<0.001
Parallelism	4	-	-	4	21	19	2	4	80	0.05	0.80

* Method of Litchfield and Wilcoxon

The expected value of each member of the series is calculated as:

		Expected number
0 colonies per plate		
n/e^m	log n : 2.2455	
	(-)4.45 x log e : 1.9326	
	0.3129	antilog = 2.06
1 colony per plate		
mn/e^m (1!)	(≠)log m : 0.6484	
	0.9613	antilog = 9.04
2 colonies per plate		
m^2n/e^m (2!)	(≠)log m : 0.6484	
	1.6097	
	(-)log 2 : 0.3010	
	1.3084	antilog = 20.36
3 colonies per plate		
m^3n/e^m (3!)	(≠)log m : 0.6484	
	1.9571	
	(-)log 3 : 0.4771	
	1.4800	antilog = 30.20
4 colonies per plate		
m^4n/e^m (4!)	(≠)log m : 0.6484	
	2.1284	
	(-)log 4 : 0.6021	
	1.5263	antilog = 33.60
5 colonies per plate		
m^5n/e^m (5!)	(≠)log m : 0.6484	
	2.1747	
	(-)log 5 : 0.6990	
	1.4757	antilog = 29.90
6 colonies per plate		
m^6n/e^m (6!)	(≠)log m : 0.6484	
	2.1241	
	(-)log 6 : 0.7782	
	1.3459	antilog = 22.18
7 colonies per plate		
m^7n/e^m (7!)	(≠)log m : 0.6484	
	1.9943	
	(-)log 7 : 0.8451	
	1.1492	antilog = 14.10
Carry forward	1.1492	

(brought forward)		1.1492	
8 colonies per plate			
m^8n/e^m (8%)	(+)log m :	0.6484	
		1.7976	
	(-)log 8 :	0.9031	
		0.8945	antilog = 7.84
9 colonies per plate			
m^9n/e^m (9%)	(+)log m :	0.6484	
		1.5429	
	(-)log 9 :	0.9542	
		0.5887	antilog = 3.88
10 colonies per plate			
$m^{10}n/e^m$ (10%)	(+)log m :	0.6484	
		1.2371	
	(-)log 10 :	1.0000	
		0.2371	antilog = 1.73
		Total expected	= 174.89

Chi-square then is calculated to determine the goodness of fit of the observed and the expected values. End members of the series are combined so that all expected values are 5 or greater and the individual contributions to Chi-square are computed as $\frac{(O - E)^2}{E}$ as shown in the last column of Table 11. Although there were 11 original observed and expected values, these were reduced to 8 by combining the end values. Two restrictions were imposed; summation and adjustment to a Poisson distribution. Accordingly Chi-square is estimated with 6 degrees of freedom. The value of Chi-square obtained 8.7541, is not significant at the 5% level, i.e., there is no reason to reject the hypothesis that the observed distribution of viable organisms follows the Poisson distribution.

The practical importance of the above is obvious when it is realized that in Poisson distributions the variance equals the mean. Thus, if the mean dose were 5 organisms, the variance of a series of doses also would be 5. The standard deviation would be $\sqrt{5} = 2.24$ and approximately 95% of individual doses would contain $5 \pm 2 \times 2.24$ organisms, or range from 0.52 to 9.48 organisms. In round numbers, certain doses would be nearly 20 times larger than others even if the suspension had been perfectly prepared.

The Binomial Distribution

The binomial distribution is of great importance in biological research since quantitative biological assays are necessary in many investigations, such as on the effectiveness ofchemotherapeutic agents or insecticides, the protective potency of active or passive immunizing agents, etc. In many instances, these quantitative estimates are obtainable only by administering progressively increased doses of the agent being studied to successive groups of experimental subjects and by determining all or nothing responses (such as survival or death) in the successive groups. Such responses commonly are designated as quantal responses.

In too many instances such quantal response data are obtained from a minimal number of groups of subjects, each group containing only a small number of individuals. It is not surprising that such data notoriously are difficult to reproduce, but it is rather surprising that experimental findings of the type being discussed continually are being published without apparent recognition of the uncertainty of the data. It is beyond the scope of this presentation to deal at length with the errors inherent to quantal response data. However, results of an extensive mouse-protection potency assay of typhoid vaccine conducted at the Army Medical Department Research and Graduate School are presented to illustrate the method of calculating the goodness of fit of a series of observed quantal responses with that expected in accordance with the binomial distribution.

In the study referred to, three different doses of typhoid vaccine A, B and C each were administered intraperitoneally to 100 female Swiss mice. The mice were distributed 5 to a jar, 20 jars for each vaccine dosage level. Six days later all mice were challenged intraperitoneally with a dose of 1000 viable S. typhosa and the subsequent deaths within 72 hours were recorded. Results of the experiment, in terms of the number of groups with 0, 1, 2, 3, 4 or 5 survivors at each vaccine dose level, are presented in Table 12.

Table 12. Observed frequency of 0, 1, 2, 3, 4 or 5 survivals in groups of 5 mice each, immunized with graded doses of typhoid vaccine and challenged with 1000 S. typhosa. Unpublished data from Army Medical Department Research and Graduate School.

Survivors per group of 5 mice	Vaccine dose (ml.)		
	A (0.006)	B (0.018)	C (0.054)
0	2	3	0
1	11	4	1
2	4	9	2
3	3	2	9
4	0	2	6
5	0	0	2

The percentage survival at each vaccine dosage level, based on the total of 100 mice in each case, were: dose A, 28%; dose B, 36%; and dose C, 66%. These percentages correspond to probabilities of survival (p) of 0.28, 0.36 and 0.66, respectively. The probabilities of death (q or 1-p) correspondingly are 0.72, 0.64 and 0.34. The expected relative frequency of groups with 0, 1, 2 —— survivors per group can be calculated by expansion of the binomial $(q + p)^k$ where k is the number of individuals per group. The relative frequencies are multiplied by n, the number of groups, to obtain the absolute number of groups expected in each case. In this problem k is 5 and n is 20.

Calculation of expected values is easily accomplished employing certain rules for computational procedure and the binomial coefficients given below.

Sample size (k)	Binomial coefficients
1	1 1
2	1 2 1
3	1 3 3 1
4	1 4 6 4 1
5	1 5 10 10 5 1
6	1 6 15 20 15 6 1
	etc

The computational procedure, as given by Snedecor (10), is illustrated with the data from the column headed vaccine dose B in Table 12. A table with 6 columns is prepared as follows:

Survivors per group	Powers of q	Powers of p	Coefficients	Relative frequency	Expected frequency
0	0.1073		1	0.1073	2.15
1	0.1677	0.3600	5	0.3019	6.04
2	0.2621	0.1296	10	0.3397	6.79
3	0.4096	0.0467	10	0.1913	3.83
4	0.6400	0.0168	5	0.0538	1.08
5		0.0060	1	0.0060	0.12
			Totals	1.0000	20.01

The value of q (in this case 0.6400) is entered in the next to the last place in the second column. This value is squared and entered into the next space above, cubed and entered into the next space above that, etc., until the column is completed. The value of p (0.3600) is entered into the second space from the top of the third column. This number is squared, cubed, etc., and the successive values entered downward until the column is completed. The appropriate binomial coefficients (1, 5, 10, 10, 5, 1) are entered into the fourth column.

Relative frequencies are obtained as the products of the second, third and fourth columns and are entered into the fifth column. For example; the relative frequency of groups with 2 survivors per group is obtained as 0.2621 x 0.1296 x 10 = 0.3397. The expected frequencies are obtained by multiplying the relative frequencies by n, the number of groups, 20 in this case.

Expected frequencies (to the nearest whole number) were calculated for groups at each dosage level and the goodness of fit of the observed and expected frequencies was tested by Chi-square in the usual manner. The results obtained were as follows:

Survivors per group	Dose A			Dose B			Dose C		
	Obs.	Exp.	X^2	Obs.	Exp.	X^2	Obs.	Exp.	X^2
0	2	4	1.00	3	2	0.50	0	0	—
1	11	8	1.13	4	6	0.67	1	1)	0.25
2	4	6)	0.12	9	7	0.57	2	3)	
3	3	2)		2	4)	0.20	9	7	0.57
4	0	0	—	2	1)		6	6	0.00
5	0	0	—	0	0	—	2	3	0.33
	$X^2[1] = 2.25$			$X^2[2] = 1.94$			$X^2[2] = 1.15$		

In the calculation of the above Chi-square values, the accepted rule of combining groups with expected values of less than 5 was not followed exactly because the number of entries would have been drastically reduced (the Chi-square values would have been even smaller than those obtained). In each column the degrees of freedom associated with Chi-square is two less than the number of values used in calculation since the restrictions of summation and fitting the binomial were imposed. In no case did the Chi-square value prove significant, i.e., the observed findings were in excellent agreement with those expected from the binomial expansion.

From consideration of the foregoing, it should be obvious that any quantitative estimate of potency based upon small numbers of animals, say 1 group of 5 at each dosage level, would be subject to extreme error as the result of sampling variation alone.

The Dosage-Mortality Relationship

In the measurement of the relationship of mortality to dosage, it usually is impossible to determine the minimal lethal dose of an agent for any specified individual. Furthermore, in any group of individuals, tolerances or susceptibilities will vary markedly, so, in addition to our inability to determine the resistance of an individual with accuracy, we are faced with the problem of sampling from a population.

For any group of animals (individuals) having varied individual degrees of resistance to an infectious or toxic agent, a certain dose of the agent can be employed which will cause the death of a certain proportion of the group. This is known as a quantal response. If groups of individuals with comparable variations in resistance are given larger or smaller doses of the agent, the proportion of deaths will be greater or smaller, respectively. By employing the proper range of doses in a series of comparable groups it then becomes possible to establish a relationship between mortality and dosage which can be expressed as a mathematical equation or portrayed as a graph. Furthermore, if the relationship of mortality to dosage of two agents is determined independently, but under comparable conditions such as in parts of the same experiment, it is possible to determine the relative effectiveness or strength (potency) of the two agents. The approach to the determination of virulence of infectious agents, protective potency of immunizing agents, or effectiveness of chemotherapeutic agents, etc., is based upon an extension of this thesis.

Obviously, the dosage-mortality relationship, as expressed by an equation or by a curve, is subject to considerable error due to sampling (e. g., resulting from binomial variation) unless the number of animals used at each dosage level is infinitely large. In addition, the relationship is exceedingly complex, i.e., mortality is not linearily related to dose. In fact, the curve of the relationship of percentage mortality to arithmetic dose is sigmoid. This sigmoid curve in general, can be made symmetrical by plotting percentage mortality against the logarithms of the doses. The general shape of these curves is shown in Figure 2.

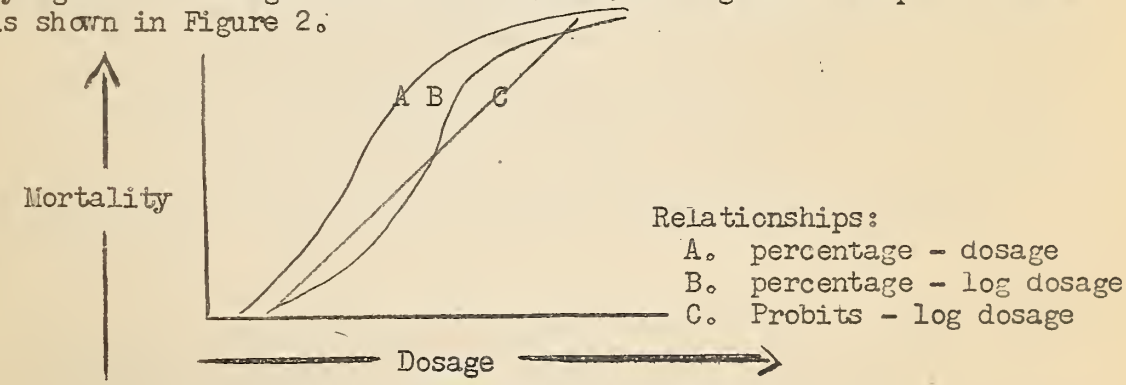


Figure 2. The relationship of mortality to dosage.

The Probit Transformation.-- For ease in computing a regression equation for predicting mortality at any dose, or for representing the relationship graphically it is of great advantage to transform the data so that the relationship is linear. Of the various transformations that have been tried, the probit-log transformation, as proposed by Bliss (12), generally is conceded to be superior. A full theoretical consideration of the probit transformation is beyond the scope of this Introduction but, in brief, the probit is the normal equivalent deviate + 5. The normal equivalent deviate is defined as

$$\text{Normal equivalent deviate } (Y) = \frac{X - \bar{X}}{s}$$

where X is the logarithm of the individual effective dose, $\bar{X}$ the mean value of individual effective doses in a normally distributed infinite population, and s the standard deviation of the population.

It is sufficient for the purpose of this discussion, and for routine calculation of dosage-mortality data, to consider that probits are a series of values corresponding to percentages and that when probits-mortality or probits-survival are plotted against log-dosage, a straight line relationship will be obtained provided reliable data are employed.

Estimation of L. $D._{50}$ and E. $D._{50}$ values.-- Whenever an assay is performed in animals to obtain some numerical expression of potency or strength of an agent, the results can be expressed in terms of either mortality or survival. It is common practice to use the term $L.D._{50}$ (lethal dose for 50%) when expressing potency of a lethal agent and to use the term $E.D._{50}$ (effective dose for 50%) when referring to a protective substance. In either case, the point of 50% response is preferred since observations at this point are subject to least error.

The most nearly exact method of determining the point of 50% response no doubt is the formal probit method based on a maximum likelihood solution. However, the method is unduly long and tedious for use except in instances where exacting estimation is required. Various methods for estimating 50% endpoints have been proposed and several of these permit estimation of such values with accuracy sufficient for most practical purposes. An especially useful and rapid method proposed by Miller and Tainter (13) is presented here. The data used for illustration of the method is that previously presented in Table 12.

The data (percentage survival) are first transformed to probits-survival by employing a table of probit values (Table V, appendix A). The probit values corresponding to percentages can be used equally well for percentage survival or percentage mortality. Values obtained are as follows:

Dose	Per cent Survival	Probit
A	28	4.42
B	36	4.64
C	66	5.41

The probit values are plotted as linear ordinate values on semi-log graph paper, the corresponding vaccine doses being plotted as absolute values on the logarithmic abscissa. With the aid of a straightedge, a line is drawn which by inspection is estimated to be the best fit to the plotted points. Not all probit values have equal weight; those near probit 5 have greatest weight and those farthest from 5 have least. In formal probit calculations weighting coefficients are employed but in this approximate method minor adjustment in the position of the line can be made to accomplish the same purpose.

Since 50% response corresponds to a probit value of 5, the E.D.50 value can be determined almost at a glance. The point where the probit 5 line intersects the regression line is noted and the corresponding value of X (dosage) is read on the base line as the E.D.50.

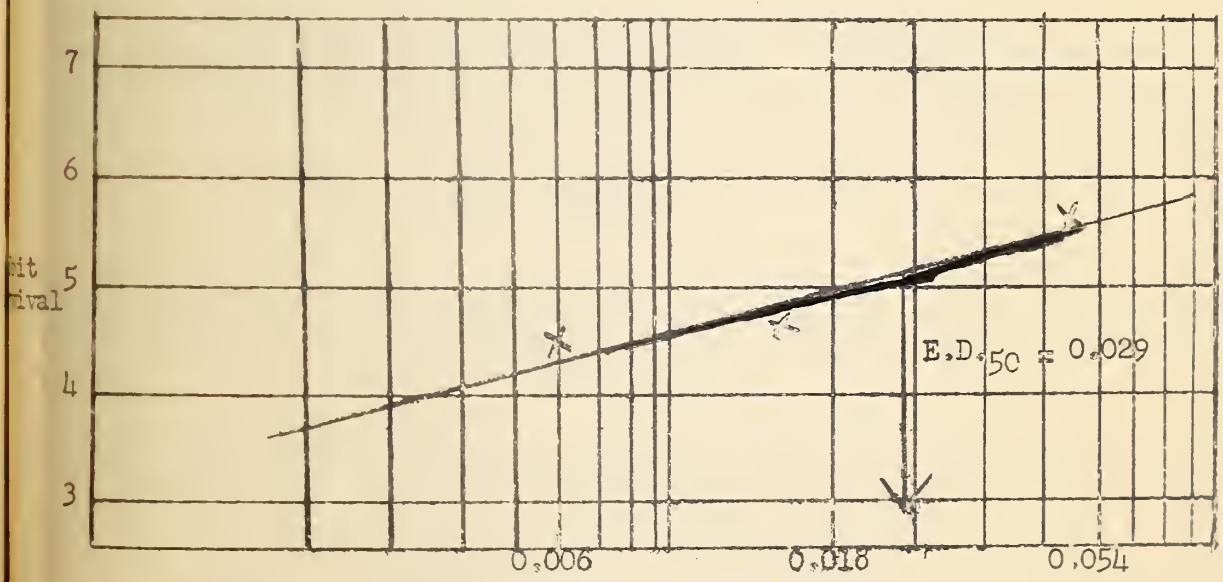


Figure 3. Graphic estimation of the E.D.50 from the data of Table 12.

Estimation of error of the $E.D._{50}$.— For accurate estimation of the standard error of the $E.D._{50}$ the Bliss method should be employed. However, an approximation of the standard error of the $E.D._{50}$ can be made by the method proposed by Miller and Tainter as follows:

$$S.E._{E.D._{50}} = \frac{2s}{\sqrt{2N}}$$

where 2 s = the difference between the values of X (mls. of vaccine) corresponding to probits 6 and 4, and N is the total number of animals used in the groups which responded in the range of 3.5 to 6.5 probits. In this problem the standard error is found as

$$S.E._{E.D._{50}} = \frac{0.200 - 0.004}{\sqrt{(2)\ (20)\ (5)\ (3)}} = \frac{0.196}{24.5} = 0.008$$

Approximate confidence limits also can be obtained as

$$\pm L = 2 \times 0.008$$

$$= 0.016$$

Then, the 95% confidence limits are 0.029 ± 0.016 or 0.013 and 0.045. Actually, confidence limits of an $E.D._{50}$ value are not symmetrical about the endpoint and for accurate results the Bliss method (14) should be used.

Determination of the relative strength of two agents.— In most practical determinations of potency, an unknown product is compared with a known or standard and its potency is expressed in terms of the standard. The relative potency of the two substances is best estimated by the following equation (14):

$$M = \bar{X}_s - \bar{X}_u \pm \frac{\bar{Y}_u - \bar{Y}_s}{b_c}$$

Where M = log-ratio of potency; $\bar{X}_s$ and $\bar{Y}_s$ = average-log-dose and average probits-response with the standard; $\bar{X}_u$ and $\bar{Y}_u$ = average log-dose and average probits-response with the unknown; and, b_c = combined slope of the two dosage-response curves.

In many actual potency tests the same doses of standard and unknown are employed. In such cases, the equation simplifies to

$$M = \frac{\bar{Y}_u - \bar{Y}_s}{b_c}$$

The value of b_c, the combined slope, can be calculated from the combined data. However, for most practical purposes an average value of the individual slopes is sufficiently accurate. The slope of the dosage-response line is

easily approximated from the graph since it is the ratio of vertical to horizontal increase. Since the horizontal line is measured in logarithmic units, the distance between any two corresponding points on consecutive cycles (10-fold in absolute units) is 1 in logarithmic units. The difference between the probit values corresponding to these points then is equal to the slope. The slope of each line is determined in this manner and the average of the two is taken. M then can be calculated directly. This is the log-ratio of potency and the relative strength is obtained by taking the antilog of M.

Bibliography

1. Wiener, A. S., Davidsohn, I., and Potter, E. L., Heredity of the RH blood types, II. J. Exper. Med., 81:63-72, 1945.

2. Greenwood, M., and Yule, G. U., The statistics of anti-typhoid and anti-cholera inoculations, and the interpretation of such statistics in general. Proc. R. Soc. Med., 8:113-194, 1915.

3. Davis, M. E., and Fugo, N. W., A simplified method for the quantitative determinations on free pregnanediol excretion in pregnancy. Proc. Soc. Exp. Biol. and Med., 66:39-42, 1947.

4. Calder, R. M., Effect of nicotinic acid on anoxia in rats. Proc. Soc. Exp. Biol. and Med., 68:642-646, 1948.

5. Keating, R. P., and Soberman, R. J., Uptake of radioiodine by thyroid glands of mice made obese by goldthioglucose. Unpublished manuscript, Army Medical Department Research and Graduate School.

6. Stefanini, M., Purification of the resin Amberlite IR-100 for blood coagulation studies. Proc. Soc. Exp. Biol. and Med., 67:22-25, 1948.

7. Godfrey, E. S., quoted by Wilson, G. S., and Miles, A. A., Topley and Wilson's Principles of Bacteriology and Immunity, Baltimore, The Williams and Wilkins Co., 3rd Edition, vol. 2, 1946.

8. Brinkhaus, K. M., Clotting defect in hemophilia: deficiency in a plasma factor required for platelet utilization. Proc. Soc. Exp. Biol. and Med., 66:117-120, 1947.

9. Tanner, F. W., Vojnovich, C., and VanLanen, J. M., Factors affecting riboflavin production by Ashbya gossypii. J. Bact., 58:737-745, 1949.

10. Snedecor, G. W., Statistical Methods, Ames, Iowa State College Press, 4th Edition, 1946.

11. Batson, H. C., Statistics in experimental immunology. Ann. New York Acad. Sc., in press.

12. Bliss, C. I., The determination of dosage-mortality curves from small numbers. Quart. J. Pharm. 11:192-216, 1938.

13. Miller, L. C., and Tainter, M. L., Estimation of the $E.D._{50}$ and its error by means of logarithmic-probit graph paper. Proc. Soc. Exp. Biol. and Med. 57:261-264, 1944.

14. Bliss, C. I., Confidence limits for biological assays. Biometrics Bull., 1:57-65, 1945.

Suggested Texts and Source Books

1. Brownlee, K. A., Industrial Experimentation, Brooklyn, Chemical Publishing Co., Inc., 1947.

2. Snedecor, G. W., Statistical Methods, Ames, Iowa State College Press, 4th Edition, 1946.

3. Fisher, R. A., Statistical Methods for Research Workers, Edinburgh, Oliver and Boyd, 10th edition, 1946.

4. Fisher, R. A., The Design of Experiments, New York, Hafner-Publishing Co., Inc., 4th edition, 1947.

5. Fisher, R. A., and Yates, F., Statistical Tables for Biological, Agricultural and Medical Research, Edinburgh, Oliver and Boyd, Ltd., 3rd edition, 1949.

6. Emmens, C. W., Principles of Biological Assay, London, Chapman and Hall, Ltd., 1948.

7. Finney, D. J., Probit Analysis, Cambridge, University Press, 1947.

APPENDIX A

ABRIDGED STATISTICAL TABLES

TABLE I

Table of Chi-square $(\chi^2)^{\#}$

Degrees of Freedom	Significance Level				
	95%	50%	5%	1%	0.1%
1	0.004	0.455	3.84	6.64	10.83
2	0.103	1.386	5.99	9.21	13.82
3	0.352	2.366	7.82	11.34	16.27
4	0.711	3.357	9.49	13.28	18.47
5	1.145	4.351	11.07	15.09	20.52
6	1.635	5.350	12.59	16.81	22.46
7	2.167	6.350	14.07	18.48	24.32
8	2.733	7.340	15.51	20.09	26.13
9	3.325	8.340	16.92	21.67	27.88
10	3.940	9.340	18.31	23.21	29.59
15	7.260	14.340	25.00	30.58	37.70
20	10.850	19.340	31.41	37.57	45.32
25	14.610	24.340	37.65	44.31	52.62
30	18.490	29.340	43.77	50.89	59.70

Abridged from Fisher, R. A., and Yates, F., "Statistical Tables for Biological, Agricultural and Medical Research", Edinburgh, Oliver and Boyd, Ltd., 1949, 3rd Edition.

TABLE II

Table of t#

Degrees of Freedom	Significance Level			
	10%	5%	1%	0.1%
1	6.31	12.71	63.66	636.62
2	2.92	4.30	9.93	31.60
3	2.35	3.18	5.84	12.94
4	2.13	2.78	4.60	8.61
5	2.02	2.57	4.03	6.86
6	1.94	2.45	3.71	5.96
7	1.90	2.37	3.50	5.41
8	1.86	2.31	3.36	5.04
9	1.83	2.26	3.25	4.78
10	1.81	2.23	3.17	4.59
15	1.75	2.13	2.95	4.07
20	1.73	2.09	2.85	3.85
25	1.71	2.06	2.79	3.73
30	1.70	2.04	2.75	3.65
	1.65	1.96	2.58	3.29

Abridged from Fisher, R. A., and Yates, F., "Statistical Tables for Biological, Agricultural and Medical Research", Edinburgh, Oliver and Boyd, Ltd., 1949, 3rd Edition.

TABLE III

Table of Variance Ratios, F#

Degrees of Freedom	N_1							
N^2	1	2	3	4	5	12	24	∞
1	164.4*	199.5	215.7	224.6	230.2	234.9	249.0	254.3
	4052.0	4999.0	5403.0	5625.0	5764.0	6106.0	6234.0	6366.0
	-	-	-	-	-	-	-	-
2	18.5	19.2	19.2	19.3	19.3	19.4	19.5	19.5
	98.5	99.0	99.2	99.3	99.3	99.4	99.5	99.5
	998.0	999.0	999.0	999.0	999.0	999.0	999.0	999.0
3	10.1	9.6	9.3	9.1	9.0	8.7	8.6	8.5
	34.1	30.8	29.5	28.7	28.2	27.1	26.6	26.1
	167.0	148.0	141.0	137.0	135.0	128.0	126.0	123.0
4	7.7	6.9	6.6	6.4	6.3	5.9	5.8	5.6
	21.2	18.0	16.7	16.0	15.5	14.4	13.9	13.5
	74.1	61.3	56.2	53.4	51.7	47.4	45.8	44.1
5	6.6	5.8	5.4	5.2	5.1	4.7	4.5	4.4
	16.3	13.3	12.1	11.4	11.0	9.9	9.5	9.0
	47.0	36.6	33.2	31.1	29.8	26.4	25.1	23.8
6	6.0	5.1	4.8	4.5	4.4	4.0	3.8	3.7
	13.7	10.9	9.8	9.2	8.8	7.7	7.3	6.9
	35.5	27.0	23.7	21.0	20.8	18.0	16.0	15.8
7	5.6	4.7	4.4	4.1	4.0	3.6	3.4	3.2
	12.3	9.6	8.5	7.9	7.5	6.5	6.1	5.7
	29.2	21.7	18.8	17.2	16.2	13.7	12.7	11.7
8	5.3	4.5	4.1	3.8	3.7	3.3	3.1	2.9
	11.3	8.7	7.6	7.0	6.6	5.7	5.3	4.9
	25.4	18.5	15.8	14.4	13.5	11.2	10.3	9.3
9	5.1	4.3	3.9	3.6	3.5	3.1	2.9	2.7
	10.6	8.0	7.0	6.4	6.1	5.1	4.7	4.3
	22.9	16.4	13.9	12.6	11.7	9.6	8.7	7.8
10	5.0	4.1	3.7	3.5	3.3	2.9	2.7	2.5
	10.0	7.6	6.6	6.0	5.6	4.7	4.3	3.9
	21.0	14.9	12.6	11.3	10.5	8.5	7.6	6.8
15	4.5	3.7	3.3	3.1	2.9	2.5	2.3	2.1
	8.7	6.4	5.4	4.9	4.6	3.7	3.3	2.9
	16.6	11.3	9.3	8.3	7.6	5.8	5.1	4.3
20	4.4	3.5	3.1	2.9	2.7	2.3	2.1	1.8
	8.1	5.9	4.9	4.4	4.1	3.2	2.9	2.4
	14.8	10.0	8.1	7.1	6.5	4.8	4.2	3.4
30	4.2	3.3	2.9	2.7	2.5	2.1	1.9	1.6
	7.6	5.4	4.5	4.0	3.7	2.8	2.5	2.0
	13.3	8.8	7.1	6.1	5.5	4.0	3.4	2.6
∞	3.8	3.0	2.6	2.4	2.2	1.8	1.5	1.0
	6.6	4.6	3.8	3.3	3.0	2.2	1.8	1.0
	10.8	6.9	5.4	5.6	4.1	2.7	2.1	1.0

Abridged from Fisher, R. A., and Yates, F., "Statistical Tables for Biological, Agricultural and Medical Research", Edinburgh, Oliver and Boyd, Ltd., 1949, 3rd Edition.

* In each cell, the three entries correspond to the 5%, 1% and 0.1% level of significance, respectively.

TABLE IV

Table of the Correlation Coefficient, r#

Degrees of Freedom	Significance level			
	10%	5%	1%	0.1%
1	.988	.997	1.000	1.000
2	.900	.950	.990	.999
3	.805	.878	.959	.992
4	.729	.811	.917	.974
5	.669	.754	.874	.951
6	.621	.707	.834	.925
7	.582	.666	.798	.898
8	.549	.632	.765	.872
9	.521	.602	.735	.847
10	.497	.576	.708	.823
15	.412	.482	.606	.725
20	.360	.423	.537	.652
25	.323	.381	.487	.597
30	.296	.349	.449	.554
100	.164	.195	.254	.321

Abridged from Fisher, R. A., and Yates, F., "Statistical Tables for Biological, Agricultural and Medical Research", Edinburgh, Oliver and Boyd, Ltd, 1949, 3rd Edition.

TABLE V

Probits or probability units for transforming the sigmoid dosage response curve to a straight line. In the body of the table is given the probit corresponding to each percentage response listed along the left edge and top.

%	0	1	2	3	4	5	6	7	8	9
	—	2.67	2.94	3.12	3.24	3.36	3.44	3.52	3.60	3.66
10	3.72	3.77	3.82	3.87	3.92	3.96	4.00	4.04	4.08	4.12
20	4.16	4.19	4.22	4.26	4.29	4.32	4.36	4.38	4.42	4.44
30	4.48	4.50	4.53	4.56	4.58	4.62	4.64	4.66	4.70	4.72
40	4.74	4.77	4.80	4.82	4.84	4.87	4.90	4.92	4.95	4.98
50	5.00	5.02	5.02	5.08	5.10	5.12	5.15	5.18	5.20	5.22
60	5.25	5.28	5.30	5.33	5.36	5.38	5.41	5.44	5.46	5.50
70	5.52	5.55	5.58	5.61	5.64	5.67	5.70	5.74	5.77	5.80
80	5.84	5.88	5.92	5.95	5.99	6.04	6.08	6.12	6.18	6.22
	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
90	6.28	6.28	6.29	6.29	6.30	6.31	6.31	6.32	6.32	6.34
91	6.34	6.34	6.35	6.36	6.36	6.37	6.38	6.38	6.39	6.40
92	6.40	6.41	6.41	6.42	6.43	6.44	6.44	6.45	6.46	6.46
93	6.48	6.48	6.49	6.49	6.50	6.51	6.52	6.53	6.53	6.54
94	6.56	6.56	6.57	6.58	6.58	6.60	6.60	6.62	6.62	6.64
95	6.64	6.66	6.66	6.68	6.68	6.70	6.70	6.72	6.72	6.74
96	6.75	6.76	6.77	6.78	6.80	6.81	6.82	6.84	6.85	6.86
97	6.88	6.90	6.91	6.92	6.94	6.96	6.98	6.99	7.01	7.03
98	7.05	7.08	7.10	7.12	7.14	7.17	7.20	7.22	7.26	7.29
99	7.32	7.36	7.40	7.46	7.51	7.58	7.65	7.74	7.88	8.09

Adapted from Bliss, C. I., "The Determination of the Dosage-Mortality Curve from Small Numbers". Quart. J. Pharm. & Pharmacol. 11:192-216, 1938